WEAPONS OF MATH DESTRUCTION

How Big Data Increases Inequality and Threatens Democracy

CATHY O'NEIL



算法霸权

数学杀伤性武器的威胁

［美］凯西·奥尼尔 著／马青玲 译

中信出版集团

版权信息

书名:算法霸权

作者:[美]凯西·奥尼尔

译者:马青玲

ISBN:9787508692067

中信出版集团制作发行

这本书献给数字难民

本书所获赞誉

见解深刻，发人深省！

——《纽约书评》

《算法霸权》对于正被广泛误用于我们生活的方方面面的数学是一种必要的批判。

——《波士顿环球报》

这本书具有启发性……奥尼尔证明，算法依赖使我们走向极端。

——美国《大西洋》杂志

《算法霸权》是一声响亮且直白的战斗号角。这本书承认，模型不会消逝：模型作为一种定位弱势群体的工具其发挥的作用非常令人震撼，但是如果用它来惩罚和剥夺公民权，那就是一个噩梦了。凯西·奥尼尔这本书之所以重要就是因为她相信数据科学。想要知道我们为什么一定要审查身边的各种模型，对这些模型提出更高的要求，就来读读这本重要的书吧。

——《弟弟》（*Little Brother*）作者、美国知名综合类博客
Boing Boing合作编辑 科利·多克托罗

许多算法服务于对权力和偏见的不平等操控。如果你不想被这些算法操控，就去读凯西·奥尼尔的《算法霸权》，解构如今在市场上横行霸道的大数据模型。

——《任何速度都是不安全的》
（*Unsafe at Any Speed*）作者
拉尔夫·纳德

下次再听到有人夸夸其谈大数据的神奇时，你就给他看《算法霸权》。他一定会受益终身的。

——美国文化新闻网站fusion资深编辑
费里克斯·萨尔蒙

帮人们找工作、找伴侣，预测模型潜移默化地塑造和控制着我们的命运。凯西·奥尼尔带我们踏上了一场愤怒与惊奇之旅，这本书的语言简单得就像是在对话，但它是重要的对话。我们对技术要小心谨慎。

——《当收入只够填饱肚子》
（*Hand to Mouth: Living in Bootstrap America*）
作者 琳达·提拉多

这本书不可或缺……尽管“数学杀伤性武器”这个话题在技术上很复杂，但是这本书能够清晰地引导读者理解这些模型系统……奥尼尔的书很优秀，展示了大数据和一个如此仰赖算法的世界的伦理道德风险……以及大数据是如何为自己、为那些对大数据是如何重新塑造我们这个世界感兴趣的人带去利益的，阅读《算法霸权》是十分必要的。

——《加拿大国家邮报》

如果你曾经也怀疑对大数据深信不疑可能存在危害，但又缺乏数学知识弄明白问题究竟出在哪儿，那么，这本书就是为你而来。

——《沙龙》网络杂志

奥尼尔是写这本书的理想人选。她是一个数学学者，后来转型成为华尔街金融工程师，又变身为数据科学家，曾参与过“占领华尔街”运动，最近还创办了算法审计公司。她是强烈呼吁限制算法影响我们的生活方式的人士之一……虽然《算法霸权》这本书里的硬知识和统计数据俯拾皆是，但这本书易于理解，甚至饶有趣味。奥尼尔的写作直接、易读，我一下午就读完了。

——《科学美国人》杂志

凯西·奥尼尔，一位转型成为数据科学家的数学理论家，传递了一个简单但重要的信息：统计模型无所不在，它们对我们日常生活许多方面的影响正日益增加……《算法霸权》提醒我们，我们生活中的一些领域已经被无形的算法掌控了。随着大数据巨头的继续扩张，来自奥尼尔的警示和提醒是受欢迎且必要的。

——《美国瞭望》（*American Prospect*）

奥尼尔自认数学迷，她的生动描述写出了大数据对我们生活的影响。

——《旧金山纪事报》

通过追踪算法在每个阶段是如何塑造人们的生活的，奥尼尔证实，我们的“机器人霸主”使用数据来区分群体，制造不公平以及干预民主选择。如果你需要和数据打交道，或者哪怕你只是大量在线数据的普通创造者之一，这都是一本必读书。

——美国知名科技博客媒体
Ars Technica

清醒，惊人，有价值……（奥尼尔）的写作清晰而准确，因为她将自己的论点瞄准了非专业读者，这使得这部分读者也能够快速阅读

这本有关算法的书。生活会受到大数据影响的人都可以读读《算法霸权》这本书——换句话说：这本书是人们用得上的。

——《亚斯本时报》

《算法霸权》通过现实生活中的案例以及生动的讲述，揭示了政府和大企业利用隐秘的算法和复杂的数学模型削弱公平、增加私权的事实。用明确打败隐秘，用明晓战胜困惑，这本书的出现能帮助我们及时改变航向。

——《人的舞台》（*The People's Platform*）作者
阿斯特拉·泰勒

凯西·奥尼尔的陈述非常好，她凭着自己的数学特长以及对社会正义的热情找出了大数据的漏洞。她提出了一个令人信服的观点，即数学模型被当作武器压榨社会边缘群体，扩大社会不公正。她的分析极好，行文引人入胜，同时，她的很多发现也可能会引起人们的不安。

——数据与社会机构创始人、《复杂》（*It's Complicated*）作者
达纳·博伊德

我虽然是一个数学家，却直到读了这本书才知道大数据武器是如何发挥作用的。尽管觉得恐惧，但这仍是一次相当愉快的阅读体验：凯西·奥尼尔眼中的算法世界既有黑色幽默，又充满愤怒情绪，就像《奇爱博士》或者《二十二条军规》。这本书能开阔读者眼界，引起读者忧思，非常重要。

——康奈尔大学教授，《X的奇幻之旅》作者
史蒂夫·斯托加茨

易读，生动……简洁而有说服力……《算法霸权》是我们这个时代的“屠宰场”……所有数据科学家以及所有相信数学模型可以取代人

类判断的企业决策者都应该阅读这本书。

——《数据和社会：点》（*Data and Society: Points*）作者
马克·凡·霍勒贝克

前言



小时候，我常常盯着车窗外的车流，研究每辆车的车牌号。我会把每个车牌号分解成素数，如：45=3×3×5。这叫作因式分解，是我最喜欢的消遣活动。我这个小数学迷对素数特别感兴趣。

我对数学的爱好逐渐发展成热爱。14岁时我参加了一次数学夏令营，带回来一个心爱的魔方。数学使我摆脱了现实世界的混乱。经过数学家们的一步步证明推导，数学不断发展，其覆盖的知识领域不断扩大。我也加入了数学领域，在大学时期主修数学，后来取得数学的博士学位。我的论文方向是代数数论，这根源于我从小就喜欢的因式分解。最后，我成为巴纳德学院的终身教授，该学院的数学系是与哥伦比亚大学联合创办的。

后来，我做了一个重大的决定：从大学离职，到顶尖对冲基金德劭集团（D.E.Shaw）担任金融工程师。我离开学术界进入金融领域，把抽象的数学理论应用到金融分析的实践中。我们所做的数据分析为一个又一个账户实现了总量达到数万亿美元的变现。起初，在新的研

究室研究全球经济让我感到既兴奋又震撼。但就在我在那儿工作了一年多的时候，2008年秋，全球金融危机爆发了。

显然，金融危机使得我曾经的庇护所——数学不仅卷入了这个世界性的问题，还助推了其中许多问题的发生。房地产危机，大型金融机构倒闭，失业率上升，在幕后运用着神奇公式的数学家们成为这些灾难的帮凶。而且，由于数学的功能特别强大（这是我热爱数学的原因之一），一旦其与科技相结合，其所造成的混乱和不幸也会成倍增长，它使得一个有着巨大缺陷的系统加速运转，进一步扩大规模，这些都是我原来不曾意识到的。

要是我们当时头脑清醒的话，就会后退一步思考，数学是怎么被我们误用的？我们该如何防止未来发生同样的灾祸？但是，金融危机发生以后，新的数学技术变得比以往更热门，其应用甚至延伸到更多的领域，每时每刻都在搅动着海量数据，其中大多数数据都是由社交媒体或者电子商务网站从使用者那里搜刮而来的。而且，数学逐渐不再关注全球金融市场动态，而是开始关注我们人类本身。数学家和统计学家一直在研究我们的欲望、行动和消费能力，一直在预测我们的信用，并用结果来评估我们作为学生、职员、情人的表现以及是否有变成罪犯的潜力。

这也就是我们所说的大数据经济，其收益前景非常可观。一个电脑程序可以在1~2秒内快速扫描成千上万份简历或是贷款申请，然后将结果整理成清晰的列表，让最有潜力的申请者位居前列。这不仅节约时间，而且公平客观。毕竟，电脑程序不像人类带有个人偏见，它只是一台处理数字的无情机器。到2010年左右，数学已深刻地介入人类事务，公众对数学这一工具的出现表示出了极大的热情。

然而，我看到的是危机。数学应用助推数据经济，但这些应用的建立是基于不可靠的人类所做的选择。有些选择无疑是出于好意，但也有许多模型把人类的偏见、误解和偏爱编入了软件系统，而这些系

统正日益在更大程度上操控着我们的生活。这些数学模型像上帝一样隐晦不明，只有该领域的最高级别的牧师，即那些数学家和计算机科学家才明白模型是如何运作的。人们对模型得出的结论毫无争议，从不上诉，即使结论是错误的或是有害的。而且，模型得出的结论往往会惩罚社会中的穷人和其他受压迫的人，而富人却因此更加富有。

我为这些有害模型提出了一个名称："数学杀伤性武器"（Weapons of Math Destruction，简写成WMD）。接下来，我将用一个例子向你们阐明这种模型的破坏性。

这个案例中的模型和很多其他的案例一样，其出发点是好的。2007年，华盛顿特区新上任的市长艾德里安·芬提下定决心对本市教学质量不佳的学校进行改革。当时，几乎每两个中学生中就有一个九年级学生是勉强毕业，只有8%的八年级学生在数学上的表现达标。为此，市长芬提设立了一个新的职位——华盛顿市教育总督，并聘用知名教育改革者李阳熙担任该职务。

当时流行的理论是：学生学得不够好是因为老师教得不好。所以，在2009年，教育总督李阳熙落实了一项旨在开除教学表现差的教师的计划。这符合当时全美教学质量差的地区所进行的改革的一种趋势，而且从系统工程学的角度看，这种想法非常有意义：评估教师。开除最差的教师，把最好的老师调到需求最紧迫的地方发挥他们的才干。用数据专家的话来说，就是"优化"学校的教师系统，尽可能保证给孩子们提供好的教育。除了那些"差"教师，谁会反对这项提议？教育总督李阳熙开发了一个叫作IMPACT的教师评估工具，至2009~2010学年末，华盛顿特区开除了评估结果垫底的2%的教师。第二学年末，又开除了5%，也就是206名教师。

华盛顿特区一所公立中学的五年级教师萨拉·韦索基似乎没有任何理由为此担心。她在麦克法兰中学仅任教了两年就得到了校长和学生

家长的一致好评。校长表扬她对学生们的教育极负责任，学生家长纷纷称她为“接触过的老师中最好的一个”。

但是在2010~2011学年末，韦索基的IMPACT评分很低。她的问题出自一个叫作增值模型的新评分系统，该系统用于评估数学教学和语言技能教学的效果。该算法给出的评分权重占她最终评分的一半，超过了学校领导和社区的评价。华盛顿特区别无选择，只好开除了她，以及另外IMPACT得分在最低限度之下的205名教师。

这看起来不完全像是一种政治迫害或者分数决定论。该学区的这一评估办法确实是有其内在逻辑的。毕竟学校领导也有可能是糟糕教师的朋友。他们可能只是喜欢这些教师的个性或是表面上的尽心尽力。糟糕教师很可能从表面看来是个好教师。所以，像许多其他的学校系统一样，华盛顿特区愿意减少人为偏差，更加注重评估得分，因为这一分数是根据实实在在的数学和阅读成绩计算得出的。华盛顿特区官员承诺，分数可以清楚地说明问题。分数更能体现公平。

韦索基当然觉得这些数字极其不公平，她想知道这些分数是怎么得来的。她后来告诉我说：“我认为没有人能理解这些分数。”一个优秀的教师怎么会得到如此低的分数呢？增值模型评估的到底是什么？

她所知道的就是，评估模型很复杂。华盛顿特区聘用麦斯迈提卡政策研究机构（Mathematica Policy Research）研发评估体系。该机构遇到的难题是测量特区学生在学业上的进步，然后计算学生的进步或退步在多大程度上归因于他们的老师。这当然不容易。研究人员知道，许多变量，包括学生的社会经济背景、是否存在学习障碍等，都会影响学生的学习成绩。评估算法必须要考虑到这些个人差异，这就是评估模型往往十分复杂的一个原因。

试图将人类行为、表现以及潜力归纳为某个算法或模型确实不是一件容易的事情。要想理解麦斯迈提卡政策研究机构处理的是什么问

题，你可以想象一个住在华盛顿特区东南部贫民区里的10岁小女孩。在一学期的学习之后，她要参加五年级的标准化测试。然后她的生活将继续下去。她可能正面对着家庭纠纷或是家庭经济困难，也许她正在搬家或是在担心她品行不良的哥哥，也许她不满意自己的体重或是在学校总被欺负。无论她在生活中经历了什么，下一学年她都要参加六年级的标准化测试。

如果你比较一下这个女孩两次测试的结果，最可能的情况是分数持平，当然更好的是分数提高了。但是如果分数下降，你能很容易地计算出她和那些优秀学生在两次测试的分数差距上差了多少。

但是，老师该为这一差距负多大的责任呢？这很难计算，而且麦斯迈提卡政策研究机构的教学评估模型只有少许数据可供比较。与之相反，像谷歌这样的大数据公司，研究人员会不断测试、监测成千上万个变量。他们可以把任一广告的字体从蓝色改为红色，将不同的版本分别投放给1000万名用户，然后追踪哪个版本获得的点击率更高，随时根据用户的反馈微调算法和操作。虽然我对谷歌公司有许多意见（接下来我将会在本书中做具体探讨），但谷歌的这种测试方法可以说是对数据的一种有效利用。

而想要计算一个人在一个学年内对另一个人的影响则复杂得多。韦索基表示："学习和教学中有太多的不确定因素，很难一一评估。"而且，试图借助对二三十个学生的考试成绩的分析评估一名教师的教学水平，从统计学上来说也是不可靠的，甚至是很可笑的。样本量太小了，一切皆会出错。如果要采用严格的统计学标准分析教师的教学效果的话，我们必须随机挑选几千个甚至数百万个学生参加考试。统计学家需要大量的数据平衡例外和反常情况。（我们在后文将会看到，数学杀伤性武器惩罚的个体往往是多数人中的例外。）

同样重要的是，统计系统需要反馈通路，以保证系统出差错时运行者能觉察到。统计学家会不断用差错训练模型，使之更加智能。若

亚马逊的推荐模型的相关性计算出错，给十几岁的女孩推荐了草坪修剪的工具书，则其网站的点击量必然会发生骤降。为此，亚马逊公司就需要不断调整模型，直到用户相关性推荐的算法运作正常为止。但是，如果没有错误反馈，大数据模型就会持续输出错误的结果，而没人试图对此加以改进。

我将要在本书中探讨的许多数学杀伤性武器都属于后者，包括华盛顿学区的教师评估增值模型。许多数学杀伤性武器都是依靠自己的内置逻辑来定义其所处理的情况，然后再以其自己的定义证明其输出结果的合理性的。这种模型会不断地自我巩固、自我发展，极具破坏力——而且在我们的日常生活中很常见。

在麦斯迈提卡政策研究机构的评分系统给予韦索基和其他205名教师差评之后，华盛顿特区开除了这些教师。但是该评分系统如何知道其决策是否正确呢？无从知道。评分系统确定这些教师是不合格者，那么别人就会认为他们是不合格者。206名“差”教师走了。仅仅是这一事实就表明了该评估增值模型的效果——该模型正在清理华盛顿特区的不合格教师。比起探索教学质量不佳的真相，评估模型所做的只不过是用分数具象化了问题。

这是数学杀伤性武器的典型反馈回路的一个示例。我们将会在本书中看到许多这样的例子。比如，当前，更多的雇主开始使用信用评分系统来评估求职者。雇主的想法是，及时支付账单的人更可能准时到岗和遵守规则。但其实，信用评分低的人中也有很多有责任感的、称职的员工。但是，雇主相信信用低和工作表现差呈正相关，这就导致了信用评分低的人很难找到工作。失业导致他们陷入贫穷，而这又进一步降低了他们的信用得分，让他们找工作难上加难。这是一个恶性循环。而雇主永远也不会知道，他们因为只关注信用评分而错过了多少个优秀的员工。数学杀伤性武器的构建过程存在着许多有害的假

设，这些模型包裹着数学精确性的外衣，流行于市场，未经检测便投入使用，而人们对此却毫无争议。

这凸显了数学杀伤性武器的另一个常见特征，即其结果往往更倾向于惩罚穷人。部分原因是数学模型是被设计来评估数量巨大的人群的。数学杀伤性武器擅长处理巨量数据，而且处理成本很低，这也是它们的优势所在。而富人通常受益于个人投入。高档律所或者大学预科学校会比快餐连锁店或者资金短缺的城市公立高中更依赖推荐和当面交流。我们在之后会经常看到这一点：特权阶级更多地与具体的人打交道，而大众则被机器操控。

没有人能给韦索基解释为什么她得了这么低的分数，这已经足够说明问题了。算法就像上帝，数学杀伤性武器的裁决就是上帝的指令。数学杀伤性武器就像一个黑盒子，其内容物是被严格保护的公司机密，如此，像麦斯迈提卡这样的顾问公司才得以收取高昂的费用。但维护算法的机密性也有另一个目的：如果被评估的人被蒙在鼓里，他们将不太可能找到系统的漏洞。他们只能努力工作，遵守规则，祈祷模型记录并回报他们的努力。但是，人们无从了解模型的具体运作方式，这意味着人们很难对模型给出的分数提出质疑或者抗议。

多年来，华盛顿的教师一直在抱怨他们遭到了评估系统武断的差评，强烈要求知道分数的由来。他们被告知这是算法的结果，很难进一步解释。很不幸，很多教师因此望而却步，不再追究，他们被数学吓到了。但有一个叫作萨拉·拜克丝的数学老师没有因此退缩，她不停地向学区领导、以前的同事詹森·卡姆拉斯问个究竟。在萨拉反复追问了几个月之后，卡姆拉斯让她等待一份即将发表的技术报告。而拜克丝回复道：“如果你自己都无法解释评估标准的根据，你怎么能保证评估的正当性呢？”但是，这就是数学杀伤性武器的本质——将问题分析的部分外包给程序员和统计师，而他们的原则通常就是，机器说了算。

即便评估模型的细节始终没有公布，萨拉·韦索基也知道，她的学生的标准化测试的分数在算法中占了很大的权重，而她对此有一些疑问。在麦克法兰中学任教的最后一学年，在开学之前，她看到她即将迎来的五年级新生在四年级期末考试中取得了惊人的好成绩。巴纳德小学29%的学生的阅读水平被评为“高级阅读水平”，这一成绩是该学区平均成绩的5倍。萨拉的很多学生都来自这个小学。

但是，开学后，她发现很多学生连简单的句子都读不好。很久之后，《华盛顿邮报》和《今日美国》的调查揭示，该学区41所学校的标准化测试试卷有大量涂擦痕迹，包括巴纳德小学。大范围纠正答案表明作弊的可能性很大，部分学校有多达70%的考场涉嫌集体作弊。

这和数学杀伤性武器有什么关系？有多方面的关系。第一，教师评估算法被视为一种可以改善教学质量的强大工具，这是开发该算法的本来目的，而在华盛顿校区，该评估算法以一种“胡萝卜加大棒”政策形式推行。教师知道如果他们的学生考试成绩不好，他们就会面临失业风险，因此他们想方设法确保学生通过考试，尤其是在经济大萧条期间劳动力市场需求疲软的时候。与此同时，如果他们的学生的表现好于其他学校的学生的话，该学校的教师和校领导将可以得到高达8000美元的年终奖金。在了解了这些强有力的激励政策的存在以及试卷被大量涂改、出现反常高分的事实之后，你就有理由怀疑巴纳德小学的四年级教师出于害怕或是贪婪修改过学生的试卷。

可以想见，如果萨拉·韦索基班级的五年级新生其上一学年的高分期末成绩是造假的，那么他们这一次真实的五年级期末成绩就会说明他们这一年的学习效果不佳，而他们的老师也会因此成为“差”教师。韦索基认为这正是她现在的遭遇。这种解释与家长、同事和校领导的观察相符，即她确实是一个好教师，而这可以帮助她澄清事实真相。

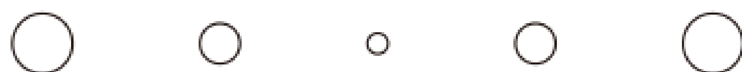
但是，你不能状告一个数学杀伤性武器。这也是我们说数学杀伤性武器具有极为可怕的破坏力的原因之一。模型不会倾听，也不会屈

服，对诱惑、威胁和哄骗以及逻辑通通充耳不闻，即使被评估者有充足的理由怀疑得出结论的数据被污染。没错，如果自动化系统出现过于明显的错误或者整体性错误，程序师的确会回头修改算法。但多数情况下，程序的裁决不容置疑，而操作程序的人只能耸耸肩，好像在说："嘿，你又能怎么样呢？"

这正是萨拉·韦索基最终得到的学校回复。詹森·卡姆拉斯后来对《华盛顿邮报》表示，试卷上的涂擦也许的确暗示了考试作弊的存在，萨拉的五年级学生前一学年的期末考分也许的确是错误的，但这些都不是决定性的证据。他声明，对韦索基老师的处理是公正的。

你看出矛盾了吗？某个算法被用于处理大量数据，它根据结果提出了一种可能性，即某人可能是糟糕的员工、有风险的借款人、恐怖主义者或者是糟糕的老师，这种可能性所对应的分数能摧毁一个人的生活。但是当有人反击的时候，作为抗衡证据的"暗示考试作弊的可能性"的涂擦痕迹又起不到作用了。之后我们将不断发现，数学杀伤性武器的受害人所面对的提供反驳证据的标准要比算法给自身设定的标准还高。

萨拉·韦索基在拿到评分结果后没几天就被解雇了。好在，很多人包括校长都担保她是个好老师，她很快在北弗吉尼亚富人区的一个学校入了职。换句话说，由于一个正当性与准确性都极为可疑的模型，穷学校失去了一个好老师，而不会根据学生考试成绩开除教师的富学校得到了一个好老师。



房地产危机发生之后，我意识到，数学杀伤性武器的应用领域已经拓展到银行业，并对整体经济造成了危害。2011年年初，我从对冲基金离职。后来我在一家电子商务创业公司担任数据分析师。因为这

一职务的关系，我发现大量数学杀伤性武器已经现身于我们能想到的任何一个行业，加剧了社会不公平，进一步压榨了弱势群体的剩余价值。这些数学杀伤性武器是正发展得如火如荼的数据经济的核心。

为了传播数学杀伤性武器这个名词，我注册了一个博客，起名叫“数学宝贝”。我的目的是动员同行数学家们反对使用草率的统计和带有偏见的模型，因为这样的统计和模型会导致恶性循环。我的博客尤其吸引数据专家，他们提醒我要将数学杀伤性武器这个概念传播到新的领域。但是2011年中期，“占领华尔街”事件在下曼哈顿区突然发酵，我意识到我们该为更广大的民众做些事情了。当时，上万民众聚集，要求经济正义和经济问责。但是当我听到记者对占领者的采访时，我发现他们似乎对经济方面的基本问题一无所知。他们明显没有读过我的博客。（这里我要多说一句，了解一个系统的缺陷，并不是要求你对整个系统都了如指掌。）

我意识到，我要么批评他们，要么加入他们，我选择了加入他们。不久后，我便推动哥伦比亚大学交替银行集团启用每周例会制度，讨论金融制度改革。在这个过程中我意识到，离开学术界之后的两次职业冒险，一次是在金融领域，另一次是在数据科学领域，给了我极大的便利接触推动了数学杀伤性武器的流行的科技和文化。

现如今，天生有缺陷的数学模型正从微观上掌控着整体经济，其影响覆盖了从广告业到监狱运营的各个领域。这些数学杀伤性武器和迫使萨拉·韦索基结束其在华盛顿特区公立中学的职业生涯的教师评估增值模型有很多相同的特点：不透明，不接受质疑，解释不通，并且都面对一定规模的大众进行筛选、定位或者“优化”。大多数数学杀伤性武器都会把其运算结果和实际情况相混淆，最终只能导致恶性循环而非问题解决。

但是，学区教师评估增值模型和用于寻找高额发薪日贷款潜在客户的数学杀伤性武器之间有一个重要的区别，即这二者会带来不同的

结果。学区得到的是一种概念上的政治货币，即教师评估得以完成，教学效果在表面上得到改善的政绩。企业得到的是本位货币：钞票。对于许多借助数学杀伤性武器运营业务的公司来说，热钱的涌入似乎证明模型奏效了。站在公司的角度，这是有意义的。当公司构建模型寻找潜在客户或者操控绝望的借款人时，越来越多的盈利似乎表明它们走对路了。但现在的问题是，利润变成了真理的象征。这种危险的混淆我们以后还会多次看到。

这种混淆的出现是因为数据科学家经常忽视交易接收端的民众。他们当然明白，数学杀伤性武器必然会出现偏差，在一段时间内会把部分人群归错类，剥夺他们找到工作或者买房的机会。但是一般来说，数学模型操作者不会思考这些可能的错误。他们看重的反馈是金钱，这也是他们的根本动机。他们设计模型就是为了吸收更多的数据，对分析结果进行微调，让更多的热钱涌入。投资者因此而尽享收益，于是决定继续将更多的钱投入数学模型开发公司。

那么受害者呢？数据科学家也许会说，没有数学模型是完美的，那些受害者是附带损失。像萨拉·韦索基这样的人常常会被他们认为没有价值，不值得惋惜。他们也许会说，别管这些人，去看那些从搜索引擎的推荐中获得有益建议的人，或是在潘多拉网络电台上找到自己喜爱的音乐的人，或者那些在领英上找到理想工作的人，还有在婚恋交友网站Match.com上找到爱情的人。多想想算法实现的这些令人惊讶的成就，忽略那些不完美。

大数据从不缺传道者，但我不在其中。本书将透视数学杀伤性武器带来的种种危害和不公正，分析其对人们在人生关键时期（如上大学，借钱，入狱，或者是找工作和保住工作）所做决策造成误导的有害例证。我们将看到，人类生活的各个方面正越来越多地被数学杀伤性武器所控制。

欢迎参观大数据的阴暗面。

第一章
盲点炸弹

不透明、规模化和毁灭性



1946年8月，一个炎热的午后，克里夫兰印第安人棒球队主帅路·波德鲁正经历着他悲惨的一天。在双重赛的第一场比赛中，泰德·威廉姆斯几乎以一人之力横扫了波德鲁的整支球队。威廉姆斯可能是当时最伟大的击球手，在这场比赛中，他粉碎了三个全垒打，为自己的球队赢得8分。最终，印第安人队以10:11遗憾输球。

波德鲁不得不采取反击。所以，当威廉姆斯在双重赛的第二场中第一次出现时，印第安人队的球员就立即开始调整各自的场地位置。游击手波德鲁换到二垒手位置，二垒手退到右外场，三垒手换到波德鲁的左边，担当游击手。很明显，波德鲁在想方设法地改变球队的防卫方向，力求截下威廉姆斯的击球。

也就是说，他在像数据科学家一样思考。他分析了原始数据——大多数都是靠观察得到的：泰德·威廉姆斯通常会把球打到右外场。然后，他据此调整了球员的站位。结果，这个策略真的奏效了。外野手接住了威廉姆斯更多的极速平直球（但依然对飞过头顶的全垒打束手无策）。

如果你今天再去看美国职业棒球联盟的比赛，你就会看到，如今，球队在制订防守策略时会把几乎每一个球员都看作威廉姆斯。波德鲁仅仅是观察了威廉姆斯通常的击球位置，而现在的球队经理则精确地知道每个球员以往击每一个球时所在的位置，包括上周的、上月的、整个职业生涯的、面对左手投手时的等等。他们利用这一历史数据分析对手的比赛策略，计算防守成功率最高的球员站位。有时候根据计算结果，全场球员都需要变换位置。

防守转移只是一个更复杂的大问题中的一个小问题。这个大问题是：棒球队可以采取哪些措施将自己获胜的可能性最大化？棒球数据科学家在寻找答案时，仔细检查了他们可以量化的每个变量，并赋予每个变量一个分值。二垒安打比一垒安打的价值高多少？什么时候值得用短打送跑垒者从一垒上二垒？

所有这些问题的答案混杂在一起，组合成了棒球运动数学模型。这些模型中的每一个都包含着各种各样的可能性，包含棒球运动要素——从四坏球、全垒球到球员素质——中所有可测量的关系。模型的目标是寻找最优组合。如果扬基队改为让右手投手应对盎格鲁队的“神鳟”麦克·卓奥特，与使用原来的投手不变动相比，扬基队有多大可能让他出局？这又将如何影响其整场比赛的胜利概率？

棒球运动特别适合建立预测性数学模型。正如迈克尔·刘易斯在其畅销书《点球成金》中所写的，棒球运动一直以来都是数据痴迷者的热门话题。过去几十年，球迷们仔细研究棒球运动员卡片背面的数据，分析C. 雅泽姆斯基的全垒球模式，或者比较罗杰·克莱门斯和杜威

·古登的出局总数。但是，从1980年开始，专业的统计学家开始分析这些数字以及大量新数据的真正意义：如何将这些数据转化为胜利，球队主理人如何用最少的钱使获胜的可能性最大化。

“点球成金”现在指针对长期被认为仅受直觉控制的领域开发的统计方法。但是，棒球模型是有益模型，与我们生活中很多领域涌现出来的有害模型即数学杀伤性武器作用相反。棒球模型之所以公平，部分原因在于其模型是透明的。每个人都可以获取作为模型根据的数据，并且或多或少能够理解模型的结果应该怎么解读。确实，一个队的模型中也许本垒击球手的表现权重更高，而另一队的模型则可能没那么看重本垒击球手的作用，因为该队的强击手经常会打出三振出局。但无论如何，在这两种情况下，全垒打和三振出局的实际次数都将展示在大家的眼皮底下。

棒球的统计也比较严谨。棒球专家手中掌握大量数据，而且几乎所有的数据都和球员的表现直接相关。可以说，他们的数据和他们根据模型预测的结果高度相关。这听起来也许平淡无奇，但读完本书我们就会看到，建立数学杀伤性武器的人通常在他们最感兴趣的行为方面缺乏相应的数据。所以，他们将本应通过调查获得的数据替换成间接变量。他们在一个人的邮政编码或语言模式和此人偿还贷款的能力或者胜任工作的潜力之间建立联系。这些联系绝大部分具有歧视性，有些甚至是不合法的。而大多数棒球模型则不使用间接变量，它们只利用最直接的相关信息，如坏球、好球和安打的次数。

最重要的是，新的棒球数据还在不断涌入，每年的4~10月，每天都有十二三场比赛的新数据涌入记录系统。统计学家可以将这些比赛结果和他们开发的模型的预测结果进行比较，以找出模型哪里出了问题。比如，他们可能预测一个左手投手会多次把击球机会让给右手击球手，但在实际的比赛中，左手投手自己击了球。这样的话，统计分析小组就得调整模型，研究哪里出错了。投球手的新曲线球会影响他

的数据吗？投球手在傍晚进行的比赛中会有更好的投地表现吗？统计学家可以把了解到的任何信息纳入模型以完善模型。这就是可靠模型的运作方式。可靠模型的开发者会对自己想要理解或者想要预测的所有事情进行反复的核实查证，并且模型必须随着具体情况的改变而改变。

棒球模型中有成千上万个不断变化的量，你也许会好奇，我们为什么能把这种模型和华盛顿特区的教师评估模型进行比较。棒球运动模型追求细节，并且不断更新；不透明的教师评估模型则似乎建立在少量的考试成绩数据之上。后者真的是模型吗？

教师评估模型确实是模型。模型只不过是某个过程的抽象表示，它可以表示棒球比赛结果、石油公司供应链、外国政府的行动或者电影院上座率。不管是电脑模型还是人脑里的模型，模型都会吸收我们知道的相关信息，并据此预测各种不同情况下的反应。我们每个人的大脑中都有成千上万个模型，这些模型告诉我们什么是我们可以期待的，并指导我们做决定。

下面是我每天使用的信息模型。作为三个孩子的母亲，家里的饭由我来做，我丈夫就不说了，他连要往煮意大利面的水里加盐都记不住。每天晚上当我开始做饭时，我的大脑就不自觉地开始分析每个人的口味。我知道我的一个儿子喜欢吃鸡肉（但是讨厌吃汉堡），另一个儿子只爱吃意大利面（最好是加一些弄碎的帕尔玛奶酪）。同时我还得考虑他们每天的口味变化，调整我头脑中的模型。显然，我的模型中有一些不可避免的不确定因素。

输入到我内在家庭饮食模型的是这些信息：我的家人的偏好，我现在有的或者我知道可以买到的食材，还有我自己的精力、时间和决心。输出的是我该如何做这顿饭以及具体做什么。我根据我的家人在吃完饭后的满意程度、他们这顿饭的饭量以及食品的健康程度来评估一顿饭做得是否成功。根据饭的受欢迎程度和被吃掉的量，我会更新

可以用于下次做饭的饮食模型。这些更新和调整让我的饮食模型成为统计学家所说的“动态模型”。

我可以很骄傲地说，这么多年以来，我已经非常擅长给家人做饭了。但是，如果我和丈夫准备外出一周，而我想给我妈妈解释我的模型，让她代替我给孩子们做饭，那我该怎么办呢？或者，如果我的那些初次为人父母的朋友想要知道我的做饭方法，那我又该怎么办呢？这时候，我就应该将我的模型具体化、形式化和系统化，也就是说使其更加数学化。如果我有野心的话，我也许可以把它做成电脑程序。

一个理想的程序将包含所有可获得的食品、食品的营养价值和成本，以及一个关于我家人口味的完整数据库：每个人对食品的好恶。但是，我很难坐下来一一列出所有的信息。我有很多关于他们争抢芦笋、不要豆角的记忆，但我很难用一个可理解的公式把它们表述出来。

较好的解决办法就是随时间发展不断地训练模型。每天输入买了什么、做了什么的相关数据，记录家里每个人的反应。我也会录入其他的参数或者约束条件。比如，我会限定只吃当季水果和蔬菜，尽量少做果酱馅饼，但不至于少到遭到家人公开反抗的程度。我还会给模型增加一些规则：这个喜欢吃肉，这个喜欢吃面包和意大利面，这个能喝很多牛奶，还总是吃什么都喜欢抹巧克力酱。

如果我把这件事当成首要工作来做的话，许多个月以后我也许就可以提出一个非常好的模型。我将把脑子里的食品管理系统、我的内部信息模型转化成一个具象化的外在模型。在建立模型的过程中，我扩大了自己对世界的影响力。我构建了一个自动化的“凯西烹饪系统”，任何人都可以操作它，即使我不在场，它也能照常工作。

但是错误总会出现，因为模型的本质就是简化。没有模型能囊括现实世界的所有复杂因素或者人类交流上的所有细微差别。有些信息

会不可避免地被遗漏。我也许会忘记在模型中加入一些规则，比如生日当天时垃圾食品的限制可以放松，或者比起用各种方法烹制出来的胡萝卜，我的家人更爱吃生胡萝卜。

因此，要建立一个模型，我们需要对各个因素的重要性进行评估，并根据我们选出的那些重要的因素将世界简化成一个容易理解的玩具，据此推断出重要的事实和行动。我们期待模型能较好地处理一种工作，同时也接受模型偶尔会像一个愚蠢的机器一样存在很多信息盲点。

有时候，这些盲点不重要。当我们在谷歌地图搜索如何去往目的地时，谷歌给出的世界模型就只有道路、隧道、桥梁，而忽略了建筑物，因为建筑物和我们想要的答案无关；当利用航空电子软件指导飞机飞行时，该软件给出的世界模型就只包含风、飞行速度和地面的着陆带，而不会显示街道、隧道、建筑物和人。

一个模型的信息盲点能够反映建模者的判断和优先级序列。谷歌地图和航空电子软件对于信息的选择似乎已经是固定不变的了，但其他模型的信息选择则存在着严重的问题。回到之前那个例子，华盛顿特区学校的教师评估增值模型主要依据学生考试成绩评价教师的教学质量，而忽视了教师对学生的投入度、在专业技能上的钻研度、教学管理方面的成果以及在帮助学生解决私人和家庭问题上的表现等。该评估模型过于简单，为追求效率牺牲了精确性和洞察力。但是，在学校领导层看来，该模型是把业绩明显不佳的几百位老师找出来的有效工具，他们并不介意这意味着模型可能会误解其中一部分人。

我们可以看出，尽管被普遍认为是不公正的，该模型还是能反映出建模者的目标和思想观念。当我在模型中排除了每餐吃果酱馅饼的可能性时，我也是在把我的思想观念强加到我的烹饪模型里。我们会毫不思索地做这件事。我们自己的价值观和欲望会影响我们的选择，

包括我们选择去搜集的数据和我们要问的问题。而模型正是用数学工具包装出来的各种主观观点。

一个模型是否奏效也见仁见智。毕竟，不管是正式模型还是非正式模型，关键要素都是其对某事成功或符合标准的定义，这一点在我们探讨数学杀伤性武器的典型特征时还会讲到。我们不仅要问是谁设计的模型，还要问设计模型的人或者组织机构要达成什么目的。比如说，如果是由某个贫困国家的政府来为我的家庭饮食建立模型，则该模型的成功可能指在我们现有食物储备的基础上，以保证我们一家不至于陷入饥饿为前提，尽可能地降低成本。个人饮食偏爱将被很少考虑或者根本不予考虑。相反，如果是由我的孩子建立模型，则成功的特征可能就是每餐都可以吃冰激凌。而我自己的模型会综合考虑资源管理和我孩子们的开心程度，还将参考我自己所确定的健康—方便—丰富—可持续性这一优先级序列。因此，我的饮食模型更为复杂。但是我的饮食模型确实反映了我的个人实际。另外，适用于今天的模型在明天的效果就不一定有那么好了。如果不经常进行更新的话，我的饮食模型就会被淘汰。食品价格会变动，家人的口味也会改变。在孩子们六岁时制定的饮食模型肯定不适用于他们进入青少年阶段的饮食习惯。

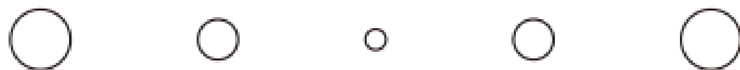
内在模型也是如此。你可能会看到这样的现象，祖父母隔了较长的一段时间再去探望孙子或孙女时往往就会出问题。上一趟来时，他们收集了孩子们知道什么、什么会让他们笑、他们喜欢什么电视节目之类的数据，然后在无意识中建立了和五岁的孩子们有关的模型。而一年后再见到孩子们，会面的最初几小时会让他们感到困惑，因为他们的模型失效了。孩子们不再觉得汤姆斯小火车头有意思了。他们需要花些时间重新收集有关孩子们的数据来调整自己的内在模型。

这并不是说，好的模型不可能简单。一些非常有效的模型可能只有一个变量。最常见的家用或办公室火灾探测模型就只测量一个与火

灾密切相关的变量：烟的出现。通常来说，这就足够了。但是当建模对象是我们的人类同胞时，只考虑简单的变量就会出问题，或者说会使我们遭遇麻烦。

种族主义在个人层面上可以被视为在全世界数十亿人的大脑中快速运转的预言模型。这种模型是基于有缺陷的、不完整的或是笼统的数据建立的。无论是来自经验还是来自传闻，这些数据都是用来表明某一类人行为恶劣的。这种模型产生了一种二元论的预测，即某一种族的所有人都行为恶劣，非该种族的人则没有这一特点。

不用说，种族主义者不会花大量时间搜集可靠数据修改他们扭曲的模型。他们的模型一旦变成一种信仰，就从此固定不变了。这种模型会生成有害假设，而且很少检测这些假设的有效性，反而满足于那些确认并巩固这些假设的数据，同时忽视反面例证。因此，种族主义是最欠考虑的预测模型，它由随机的数据采集和假性相关所驱动，被制度不公平加以强化，又被证实性偏见加以进一步劣化。这么说来，种族主义和我要在本书里探讨的所有数学杀伤性武器十分相似。



1997年，非裔美国人杜安·巴克，一个已被定罪的杀人犯，在得克萨斯州哈里斯县法庭接受审判。巴克已被证实杀了两个人，陪审团必须要决定最后的裁决是死刑还是终身监禁、保留假释机会。检察官一方奋力争取死刑判决，理由是巴克如果被释放还会再杀人。

巴克的辩护律师带来了一个专家证人，心理学家瓦特·基哈诺，不过对于巴克，他一点儿忙也没帮上。基哈诺曾经研究过得克萨斯州监狱系统的累犯率，在法庭上，他提及巴克的种族与累犯率的相关性。在盘问证人时，检察官抓住了这一点。

“你断定，种族因素，黑色人种，会因为各种各样的原因带来社会上危险行为的增多。对吗？”检察官问道。

“是的，”基哈诺回答说。于是，检察官在做案件总结陈述时强调了这一证词。最终陪审团判定巴克死刑。

三年后，得克萨斯州检察长约翰·康奈发现，上面那位心理学家在另外6个死刑案件中给出了同样的种族论证词，大多数案件发生于他在检察机关工作期间。康奈——其后来于2002年当选美国参议院议员——下令为涉案的7名犯人重新召开不带有种族偏见的听证会。在媒体发布会上他声明：“刑事司法体系中，将种族因素纳入考虑范畴是不合理的……得克萨斯州人希望，也值得拥有人人平等的司法体系。”

这7名犯人中的6名重新接受了审判，但他们再次被判处死刑。法庭裁决，基哈诺的带有偏见的证词不是决定性因素。巴克没有得到重新审判的机会，也许是因为提出种族论证词的己方证人。他仍是死刑犯。

不管在审判时包含种族因素的证词是否被明确提出，很长一段时间里，种族都是影响审判结果的一个主要因素。马里兰大学的一项研究表明，在哈里斯县，包括休斯敦市，对于犯下同等罪行的犯人，检察官判非裔美国人死刑的概率比白人高3倍，判拉美裔美国人死刑的概率比白人高4倍。这种情况并不是得州独有的。美国公民权利联盟的调查显示，犯同样的罪，黑人罪犯的刑期比白人罪犯的长20%。黑人只占据美国总人口的13%，但黑人罪犯占据了美国40%的牢房。

你可能会认为，利用电子化、数据化的再犯风险模型辅助判决能减少偏见对判刑的影响，更有利于实现公正判决。美国24个州的法院正寄希望于此，于是其采用了所谓的再犯模型作为辅助工具。再犯模型被用于帮助法官评估每一个罪犯的危险性。从很多方面来说，再犯模型的开发是一种进步，它使得审判更具一致性，更少被法官的情绪

和偏见所影响。另外，再犯模型减少了罪犯的平均刑期，节省了政府开支。（关押一个犯人一年平均需花费31000美元，在康涅狄格和纽约州，该项成本还要翻一倍。）

但问题是，我们是彻底根除了人类偏见，还是只不过用技术包装了人类偏见？再犯模型的开发是一个非常复杂的数学问题，而再犯模型的框架是由大量的假设构成的，其中一些假设本身就带有偏见。而且，瓦特·基哈诺的公开证词在被转录成文字之后，还可以供他人在法庭上阅读和质疑，但一个再犯模型的运作完全是由算法独立完成的，只有极少数专业人士能理解。

一个更普及的用于评估罪犯危险性的模型，叫作LSI–R（水平评估量表），其中包含一个需要罪犯填写的冗长的问卷。其中一个问题，“你之前被定罪过几次”与再犯风险高度相关。其他问题也非常相关，比如“其他人对你这次犯罪起了多大的作用？”“毒品和酒精对你这次犯罪起了多大作用？”

但是，当问题延伸到深挖罪犯的个人生活时，我们很容易想到，有特权背景的罪犯和来自治安差的城市贫民区的罪犯，他们的答案肯定不一样。问一个在舒适郊区长大的罪犯“你第一次遭遇警察”的原因，他也许会告诉你这次入狱就是第一次。相反，生活在贫民区的年轻黑人男性很可能已经被警察拦截过许多次了，即使他们什么错事也没做。纽约公民权利联盟2013年发表的一份研究报告显示，14~24岁的黑人男性和拉丁美洲男性仅占该市总人口的4.7%，但其占被警察“拦截—盘查”总人数的40.6%。在这些被盘查的少数族裔中，超过90%的人都是无辜的，还有一些也许只是犯了未成年酗酒或者携带大麻的轻罪。不像大多数富人孩子，他们总会因为这些小事遭遇麻烦。所以，如果曾在早期“遭遇”过几次警察就表示一个犯人是惯犯，这对穷人和少数族裔是很不公平的。

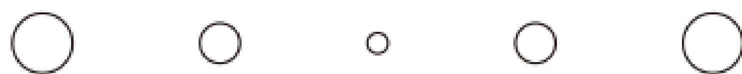
该问卷还没有结束。罪犯还会被问及他们的朋友和亲戚是否有过犯罪记录。同样，问在中产阶级社区长大的罪犯这个问题，得到否定回答的可能性很高。调查问卷确实回避了种族问题，因为问种族问题是非法的，但是有了每个罪犯提供的大量生活背景细节，这个非法的问题也没必要再问了。

自1995年LSI–R调查问卷投入使用以来，已经有成千上万个罪犯做过这张问卷了。统计师利用所收集的答案设计出了一个模型，其中与再犯率高度相关的问题答案权重更高。罪犯在答完调查问卷之后，模型会基于他们的分数将其划分为高、中、低三种风险等级。在美国的有些州，比如说罗得岛州，这一测试仅用于找出那些正被监禁的罪犯中风险等级高的人，将其送入强化的劳改项目。但是在其他州，包括爱达荷州和科罗拉多州，法官会用模型给出的评分指导量刑。

这是不公平的。这份调查问卷涉及罪犯的出生地和成长环境，还包括他的家庭、所在街区和朋友，而这些细节不应该被视为和刑事案件或者量刑存在相关性。如果检察官企图通过提及被告兄弟的犯罪记录或者其所在街区的高犯罪率去判定被告的话，正义的辩护律师就会大喊："法官大人，我反对！"而严肃的法官会判定反对有效。这是我们法律系统建立的基础。我们应该因为我们所做的事情而接受相应的审判，而不应该因为我们的身份而被审判。虽然我们不知道这些问题在问卷中所占的确切比重，但可以肯定地说，任何大于零的比重都是不合理的。

很多人会说，像LSI–R这样的数据模型有助于评估罪犯的再犯风险，或者说至少比法官的随意猜测要更精确一些。但是，即使我们暂且不谈重要的公平问题，我们也已经陷入数学杀伤性武器创造的恶性循环之中了。得到"高风险"评分等级的人很可能本来就是失业人员，在其所生活的社区里，他的许多朋友和家人都触犯过法律。得到这一评级是导致其刑期变长的一个原因，而多年和一群罪犯关在一起又增

加了他再次犯罪的可能性。等他出狱之后，他又会回到同样的贫穷社区，而这一次还有了犯罪记录，对他而言，找工作变得更难了。如果他因生活所迫不得不再次犯罪，再犯模型就又一次得到了成功验证。但事实上，正是这一模型本身导致了犯人陷入恶性循环，并且进一步巩固了犯人的恶劣处境。这是数学杀伤性武器的典型特点。



这一章，我们已经研究了三种模型。棒球模型基本上是一种健康模型。这种模型信息透明，不断更新，假设和结论大家都可以看到。棒球模型仰赖比赛进行过程中积累的真实数据，而不是替代性的间接变量。而且模型涉及的球员都明白比赛过程，且和模型的目标一致：赢得世界职业棒球大赛。（当然，这并不是说合同期内的球员不会对模型的评估结果发牢骚："没错，我确实出局200次，但是请看看我的全垒打……"）

就我个人而言，我们讨论的第二种模型，家庭饮食模型，绝对是一种良性模型。如果我的孩子们要质疑模型涉及的某个假设，不管是经济上的还是饮食上的，我都会很乐意回答他们。即使有时候他们看到盘子里的绿色蔬菜会摆臭脸，但他们仍然会承认，大家在家庭饮食上的共同目标是方便、省钱、健康、美味，只不过在每个人自己的模型里，各要素的分量有所不同。（当他们开始自己做饭之后，他们就可以建立自己的模型了。）

我要补充说明的是，我的饮食模型绝对不可能规模化。我并不乐于看到沃尔玛、美国农业部或其他任何大型机构拥护我的模型，并强行将其施加到亿万人的生活中，就像应用那些我们在本书中要讨论的数学杀伤性武器一样。不，之所以说我的饮食模型是良性的，极其重要的一点是因为我的饮食模型永远不会离开我的大脑，不会变成一串固定的代码。

但是，本章最后的再犯模型则与前两者完全不同。让我们迅速做一个简单的数学杀伤性武器判定练习，看看它是否属于此类模型。

第一个问题：如果参与者知道自己是被模型评估的一个对象，或者知道模型的目的是什么，那么该模型还是不透明，甚至是隐形的吗？绝大多数填写强制调查问卷的罪犯都不是蠢蛋。他们多少都会怀疑自己提供的信息将被用来安排自己的监狱生活，比如会被关押更长的时间或更短的时间。他们知道游戏规则。但是监狱官也知道。因此，他们对LSI–R调查问卷的目的只字不提。否则的话，他们知道很多罪犯会弄虚作假，在离开监狱的那天做再犯风险调查问卷时回答得像个模范市民。所以，罪犯需要被尽可能地蒙在鼓里，不被告知自己的风险等级评分。

再犯模型远非个例。不透明、隐形成了这类模型的规则，清晰、透明的模型倒成了例外。我们被模型分类为购物者、沙发懒虫、病人和贷款申请者，而我们自己对此知之甚少，甚至仍在愉快地注册各种把我们当成评估对象的应用程序。即使这些模型是良性模型，不透明还是给人一种不公平的感觉。如果你在进入一个露天音乐会现场之后，导引员跟你说你不能坐在前十排，你会觉得这很不合理。但是如果导引员跟你解释前十排是为行动不便的人保留的，那你的感觉就大不一样了。所以，透明很重要。

然而现实是，许多公司竭尽所能地隐藏它们的模型运算结果，甚至隐藏模型的存在。常见的一个辩护理由就是模型算法包含对它们的业务至关重要的“商业机密”。这是知识产权，如果有必要，公司必须在大批律师和说客的协助下为其维护算法机密性的行为进行辩护。比如谷歌、亚马逊和脸书这样的互联网巨头，它们为自己的业务量身定做的算法价值高达数十亿美元。数学杀伤性武器是个深不可测的黑盒。因此，明确回答第二个问题特别困难：模型违反国民主体的利益

吗？简单来说，模型是不是不公平的？它会破坏或毁灭一些人的生活吗？

根据对于这个问题的回答，LSI–R再一次成为数学杀伤性武器的典型。毫无疑问，20世纪90年代建立该模型的人认为，LSI–R是提高刑事司法系统的公平和效率的一个有效工具。它能帮助没有威胁性的罪犯缩短刑期，而这部分罪犯将因此获得更多年的自由时间，同时这也将大大节省美国纳税人的钱，毕竟每年用于监狱运营与管理的财政开支高达700亿美元。但是，再犯风险调查问卷是根据犯人的生活背景细节信息评判罪犯的危险等级的，而该细节信息在法庭上是不被允许作为证据出现的，因此这个模型是不公平的。虽然很多人可能会因此受益，但另一些人也因此受苦。

导致一部分人受苦的关键原因是模型造成的恶性循环。我们看到，再犯模型会根据一个人的成长环境来描述这个人的基本情况，它会自行创建一种使假设合理化的环境。而模型则在此恶性循环的过程中变得越来越不公平。

第三个问题：该模型是否有应用场景呈指数增长的潜力？用统计学家的话来说就是，该模型能否规模化？这听起来可能像是一个书呆子数学家的较真，但是规模化的确增强了数学杀伤性武器的破坏力，使其逐步转变为我们生活中的决定性因素。我们将会看到，不断发展的数学杀伤性武器在人力资源、健康、银行等数不尽的行业快速确立普适准则，继而对我们产生一种非常类似于法律的权威性影响。比如，如果你被银行的模型认定为高风险贷款者，那么所有人都会把你当成赖账不还的人，即使你完全不是这样的人。当这个银行的模型规模化后，就像现在的信贷模型那样，你的一生都将生活在其阴影下，你能否买到公寓、找到工作或者买到车等，都将由这一模型来决定。

就规模化而言，再犯模型再次成为一个典型。大多数州已经投入使用这一评估模型，而LSI–R是其中最常见的一个，至少已在24个州

中投入使用了。罪犯为数据科学家提供了一整个活跃的市场。刑罚体系积累了大量数据，因为罪犯比平常人享有更少的隐私权。而且，刑罚体系因为太过于臃肿、低效、高成本、缺乏人性而亟待改进。谁不想要这样一个低成本的模型应用场景呢?

刑罚改革在今天这样一个极化政治世界是一个极为罕见的议题，自由党和保守党在这一议题上有着共同的利益。2015年年初，保守党的科氏兄弟，查尔斯和大卫，与自由党的智库“美国进步中心”合作推进监狱改革，致力于减少监狱人数。但是，我对两党合作加上其他一些团队的共同努力，是否一定能够提高用于监狱的评估模型的效率和公平依然持怀疑态度。即使其他的工具取代LSI–R成为监狱中的主要评估模型，监狱系统仍然是大规模数学杀伤性武器的强大孵化器。

综上所述，数学杀伤性武器共有三个特征：不透明、规模化、毁灭性。这三个特征将在我们之后讨论的案例中有不同程度的呈现。是的，确实存在有争议的地方。比如，你可能会争辩说再犯模型的评级分数不是完全不透明的，罪犯在一些情况下是可以看到自己的评级的。但是这个模型还是太神秘了，罪犯不知道根据他们给出的问卷答案，模型是如何推导出他们的分数的。评分算法是隐藏的。另外，少数数学杀伤性武器似乎不满足规模化的特征，因为它们的规模还不够大，或者至少现在还不会规模化。但它们是蓄势待发的危险物种，可能会在将来的某一天突然开始以指数级的增长速度繁殖。所以我也把它们算在内了。最后，你可能还会指出，并不是所有的数学杀伤性武器都是有害的。毕竟，有些模型把一部分人送进了哈佛，让一些人得到了低息贷款或者找到了工作，缩短了某些幸运重刑犯的刑期。但重点不是有没有人受益，而是有很多人受害。这些数学杀伤性武器关闭了亿万人的机会之门，通常只是因为一些微不足道的理由，而且不予他们上诉的机会。因此，它们仍然是不公平的模型。

还有一个关于算法的事实是：算法能从一个领域跳跃性地应用于另一个领域，而且经常如此。传染病学研究中的模型被用于预测票房，垃圾邮件过滤器的模型被用于发现艾滋病病毒。数学杀伤性武器也是如此。所以，如果监狱中的评估模型得到了成功应用（其实际功劳完全可以归结为人类的有效管理），则它也将像其他的数学杀伤性武器一样延伸到其他领域，给我们带来附带伤害。

以上就是我的观点。威胁还在扩大。对此，金融领域已经为我们提供了一个警世寓言。

第二章
操纵与恐吓

弹震症患者的醒悟



假如你有个习惯，每天早晨，在赶从朱利叶市到芝加哥拉萨尔街的列车之前，你都要塞两美元到咖啡自动贩售机里，而咖啡机会找回你两枚硬币，并“吐”出一杯咖啡。但有一天，咖啡机找回你4枚硬币。下个月，同样的机器又出现了三次同样的情况。那么这就说明，一种模式正在形成当中。

如果这是出现在金融市场中的小反常，那么像我这样的曾就职于对冲基金的金融工程师可能就会注意到这件事。金融工程师会研究几年甚至几十年的数据，然后建立一种算法来预测这个循环出现差错——多找回的两枚硬币——的可能性并对此下注。即使只是一个规模很小的模式，也能为第一个发掘该模式的投资者带来数百万元美元的收益。这些模式不断产出大量盈利，直到发生以下任意一种情况：盈利模式消失了，或者市场中的其他人也了解了这个模式，先行者优势

消失。到了那个时候，好的金融工程师将会去追踪其他的盈利性的市场反常模式。

探索市场失灵现象就像寻宝游戏，很有意思。当我逐渐习惯于自己在德劭集团的新工作时，我发现退出学术界是个可喜的转变。虽然我喜欢在巴纳德学院任教，也热爱我的代数数论研究，但我发现数学研究的进度实在是太慢了。我希望成为这个快节奏的世界中的一分子。

当时，我认为对冲基金在最坏情况下也只不过是道德中立的金融系统的清道夫。德劭集团算是对冲基金界的哈佛，能在这样的公司工作，并向他人展示我的智慧可以变现，我很骄傲。我的工资增长到了当大学老师时的工资的三倍。刚开始这份新工作的时候，我从未想过自己会站在金融危机的最前线，而数学在其中又发挥了怎样的破坏力。在对冲基金工作，让我得以近距离研究数学杀伤性武器。

起初，我有很多喜欢做的事情。德劭集团的所有业务都是建基于数学之上的。在很多证券公司，交易员操控全局，做大交易，厉声下命令，赚取数百万美元的红利，而金融工程师只是他们的部下。但是在德劭集团，交易员只是普通的行政职员，他们被称作执行者，真正掌握全局的是金融工程师。我所在的十人团队是“期货组”（futures group）。在一个一切都取决于明天发生什么的行业，还有什么能比未来更重要呢?

我们公司大约有50个金融工程师。早期，除了我之外，其他人全是男性，而且绝大多数都不是美国人。很多人都来自理论数学或者理论物理领域；少部分人，比如我，来自数论领域。但是，我没有太多机会和他们进行专业上的交流。因为我们提供的想法和算法是对冲基金的业务根基，所以很显然我们这些金融工程师也是一个风险因素：如果我们离职了，我们可以用我们的知识快速为原东家打造一个凶猛的竞争对手。

为了防止这种风险大范围发生对公司造成威胁，德劭集团几乎完全禁止我们和其他组的同事讨论业务，有时候甚至和自己办公室的同事讨论都是不被允许的。从某种意义上来说，信息被隔绝在一个个网络细胞单元里，和基地组织也没什么差别。这一制度的作用是，如果一个细胞崩塌，换句话说，如果我们中有人突然投奔桥水基金或者摩根集团，或者自己创立公司，我们只能带走自己的知识和信息，而不至于影响德劭集团其他业务的正常运转。可以想象，这种做法将保证整个团队的士气不受影响。

期货组每一个新来的同事在入职后的前13周都被要求随时待命。这意味着，从亚洲市场开市起（美国时间是星期天晚上），到下一周的周五下午4点纽约市场收市，期间新同事要随叫随到以便处理各种问题。睡眠剥夺只是副作用之一，更糟糕的是在解决问题时因为信息不共享的公司制度而形成的无力感。比如，某个模型的算法可能出了问题。此时，我必须要找到出问题的那部分算法具体是什么，然后再去找它的负责人。而无论白天黑夜，相关负责人都要过来解决问题。因为这种麻烦事而打扰别人的情况时有发生，因此同事间互相碰面的情景总是不那么愉快。

还有各种慌乱的时刻。假期期间，很少有人工作，而麻烦事往往会在此时发生。在我们巨大的投资组合里会有各种金融产品，包括远期外汇交易，也就是承诺在未来几天大量购买外汇。我们并不需要客户真的付钱完成交易，而是会每天将交易日推迟一天，这样我们就能继续进行对于市场涨跌的判断，同时我们也不需要拿出大量的现金。在某个圣诞节假期，我发现一个购买大笔日元的交易的交易日快要到了，正常而言，此时必须有人去推迟交易日。这项工作以往是欧洲一个同事负责处理的，但他正和家人一起过圣诞节。如果没有人及时处理这个问题的话，理论上来说，我的同事就必须亲自携带5000万美元前往东京购买日元。摆平这个问题花了我很大的力气，也彻底毁了我的圣诞假期。

如果说这些问题只属于职业性危害，那么真正的问题更让人感到恶心。我已经习惯于在海量货币、债券、股票以及国际市场中的数万亿美元里进行各种操作。但是，与我的学术模型里的数字不同，商业模型里的数字是有实际意义的，它们是人们的退休基金和抵押贷款。现在回想看看，这似乎是一个极为显而易见的事实。当然，我一直知道这个事实，但从前我并没有真正领会其本质。我们利用数学工具搜刮来的每一分钱，它们不是从矿山里面挖出的金砖或者从西班牙沉船里捞出的硬币那种意外财富，它们是人们辛苦挣来的钱。而在华尔街那些自鸣得意的对冲基金投资人那里，这些血汗钱被称为“傻瓜资金”。

2008年市场崩盘时，这一丑陋的事实彻底暴露了。而比从民众的账户里窃取傻瓜资金更卑劣的是，金融行业正在致力于开发更有效的窃取资金的数学杀伤性武器，我也是其中的一分子。

其实麻烦早在2007年就出现了。2007年7月，银行间同业拆借利率上调。在2001年“9·11”恐怖袭击事件导致的市场萧条发生之后，低利率助推了房地产市场的繁荣。似乎任何人都能轻易申请到抵押贷款，房地产建筑商大规模开发郊区、沙漠和草原，银行把大量资金投入到了和房地产热潮有关的各种金融工具上。

而利率上调发出了危机即将到来的信号。银行间不再信任彼此有能力偿还隔夜贷款。它们开始慢慢地处理自己的投资组合中的“危险垃圾”，并明智地做出判断：其他银行也正面临着，即使不是更多，也是同样多的风险。现在回头看，你可能会说，利率上调其实是市场回归理智的一个信号，尽管已经太迟了。

对于德劭集团，这些前期的波动影响很小。的确，许多公司陷入水深火热之中。银行业即将遭受冲击，而且很可能是特别大的冲击。但这不是我们的问题，我们从不草率地投入风险市场。毕竟，对冲基金的本质就是对冲风险。早期，我们称这次市场波动为“混乱”。对于

德劭集团而言，这场混乱可能会造成一些不适，或者带来一两个不愉快的小插曲，但这都不是什么大问题，就像一个有钱男人在高档酒店刷不了自己的信用卡一样，只是一点点小尴尬而已。

毕竟，对冲基金并没有制造这些市场，只是参与其中。这意味着，当市场崩盘时，市场废墟会产生大量的机会。对冲基金并不是要驾驭市场，而是要预测市场的动向。市场下跌也同样有钱可赚。

想象一下芝加哥瑞格利球场的世界职业棒球大赛。由于在第9局的最后打出了一个戏剧性的全垒打，芝加哥小熊队赢得了1908年以来的第一个世锦赛冠军奖杯，他们上一次夺冠还是在西奥多·罗斯福总统任职期间。体育场的绝大部分观众欢呼雀跃，只有几排球迷坐在那儿安静地分析比赛数据。这些人不像传统的赌球者只赌比赛输赢。相反，他们会赌扬基队的投手会更多地打出三振出局而非更多地上垒，会赌这场比赛可能至少会出现一个短打，但是不会超过两个短打，或者赌芝加哥小熊队首发球员的在场时间将持续至少六局。他们甚至会赌其他的下注者是赢还是输。这些人押注赛事的众多相关方面，但不押注比赛结果本身。他们所进行的赌注就类似于对冲基金的运作模式。

这种做法让投资人感到安全，或者至少比不进行对冲操作的投资安全。我记得在某个盛大的公司活动上，公司邀请了美联储前任主席艾伦·格林斯潘和克林顿总统任职时期的财政部部长、高盛集团董事罗伯特·鲁宾。1999年，鲁宾参与推翻了于20世纪30年代“大萧条”时期美国国会通过的《格拉斯–斯蒂格法案》，致力于拆除银行业和投资业中间的屏障，这带来了接下来近十年商业、股市投机买卖的盛行。银行可以自由贷款（很多是欺诈性贷款），以证券的形式将贷款卖给客户。这看起来很正常，而且可以看作银行为客户提供的一项服务。但是，由于《格拉斯–斯蒂格法案》已经被推翻，银行现在可以对赌卖给客户的债券了。这为客户带来了巨大的风险，也为对冲基金创造了无

尽的投资机会。毕竟，我们赌的是整体市场走势，而那几年里，整个市场疯狂上涨。

在那次盛会上，格林斯潘警告过我们资产抵押债券的问题。后来我总是回想起他的警告，因为我意识到，几年之后被任命为花旗银行董事长的鲁宾在银行的大量投资组合中纳入了大规模的此类债券，这也成为后来花旗银行陷入巨额亏损，不得不求助于政府用纳税人的钱为其纾困的主要原因。

坐在他们旁边的是鲁宾的部下，也是我们公司的兼职合伙人拉里·萨默斯。他以前在财政部是鲁宾的部下，后来出任哈佛大学的校长。但是萨默斯和学校教师的关系并不好。教授们对其言论提出了集体抗议，拉里称，数学和自然科学专业的女学生数量较少，是遗传基因所致，他称之为“天资”的不公平分配。

萨默斯从哈佛大学离职以后就来了德劭集团。我记得德劭的创始人大卫·肖开玩笑说，萨默斯从哈佛离职来到德劭是“升职了”。尽管市场可能正处在震荡时期，但是德劭集团依然风生水起。

然而，随着金融危机加剧，德劭集团的合伙人开始担忧了。毕竟，所有的市场都是紧密联系在一起的——大家都是一条船上的人。当时，业内盛传雷曼兄弟已处在破产边缘，而雷曼兄弟持有德劭集团20%的股份，我们的很多交易都是交由雷曼兄弟处理的。由于市场持续动荡，德劭集团内部开始人心浮动。是，我们仍然可以用强大的模型处理大量的数据，但是，如果可怕的明天和之前的每一天都不一样了怎么办？要是明天的一切都变了呢？

这的确值得忧虑，因为数学模型的本质是基于过去的数据推测未来，其基本假设是：模式会重复。不久之后，公司的“股票组”团队以极大的代价抛售了大量股票。一直持续的金融工程师招募热潮也结束了。虽然人们仍在试图对这些新变化一笑了之，但是恐惧在增加。所

有人都在盯着证券产品，尤其是格林斯潘警告过我们的资产抵押债券。

几十年来，抵押证券一直很安全。抵押债券是一种金融工具，个人和投资基金利用资产抵押债券增加其投资组合的多样性，其原理就是增加投资品种的数量可以抵消风险。每一笔房屋抵押贷款都有违约的风险：比如房主宣布破产，这意味着银行再也收不回贷出去的全部资金；另一种极端情况是贷款人提前偿还抵押贷款，终止付息。

所以，在20世纪80年代，投资银行家开始购买成千上万的房屋抵押贷款，并将其打包成一种债券，也即一种为客户按季度定期支付股息的金融工具。可以肯定有少数房主会违约，但是，大多数人还是会持续支付房屋抵押贷款，创造稳定并且可预期的现金流。随着时间的推移，这些债券发展成一个产业，成为资本市场的支柱。专家把抵押贷款分成不同的类别或组别，一些被认为很安全，另外一些则更具风险性，相应的利率也就更高。投资者有理由充满自信，因为标准普尔、穆迪和惠誉等信用评级机构已经研究过这种债券，并给出了相应的风险评级。这些评级机构认为购买抵押债券是明智的投资。但值得注意的是，这种债券并不是透明的。投资者对于债券中抵押贷款的质量是不明的，他们对其唯一的一点了解就是信用评级机构给出的评级。而这些信用评价机构给出的评级实际上并不诚实，它们会向有金融产品要评级的公司收取相应费用而给出虚高的评级，其中，资产抵押债券无疑是一个理想的诈骗平台。

如果你想要我给你打个比方，针对这一债券，最常见的一个比喻就是香肠。你可以把房屋抵押贷款想成一片片质量参差不齐的肉，把资产抵押证券想成一捆捆的香肠——把所有东西搅拌在一起，外加一些刺激性的调味料，香肠就做成了。显然，香肠质量不一，而且从外表很难看出香肠的原料如何，但是因为美国农业部说这是安全食品，我们便毫不担心地吃下去了。

后来，人们了解到，在市场繁荣时期，抵押贷款公司通过给买不起房子的人们发放贷款赚取了巨额利润。策略很简单，就是与客户签订一张不可连续签订的抵押贷款合约，收取大额费用，然后将其包装成证券（制作成香肠）投放到繁荣的抵押证券市场。有一个臭名昭著的案例是，草莓采摘工人阿尔贝托·拉米雷斯每年的收入约在1.4万美元，他通过贷款在加州大牧场地区买了一套售价72万美元的房子。他的房产经纪人明确告诉他，他几个月以后就可以重新贷款，而且可以通过炒房赚取大额差价。而几个月以后，他就陷入了拖欠贷款的麻烦。

在房地产市场崩溃的前夕，抵押贷款银行不仅频繁地提供这种不可持续的交易，而且还积极地在贫困社区和少数族裔聚居区中寻找受害者。在一个联邦诉讼案件中，巴尔的摩市的官员指控富国银行针对黑人社区进行所谓的“贫民区贷款”。据负责处理贷款业务的前银行工作人员贝丝·雅各布森透露，富国银行的“新兴市场”部门专注于面向黑人教堂拓展业务，即让受人信任的牧师引导其教民向银行贷款。这些贷款都是次级贷款，利率最高。富国银行甚至会用这样的方式将钱贷给那些信用可靠的借款人，而后者本应该有资格获得更优惠的贷款。到2009年巴尔的摩市的官员提起诉讼的时候，有超过一半丧失赎回权的房产被富国银行收回，其中71%的房产位于非裔居民聚集区。（2012年，富国银行解决了诉讼，同意向全美各地的3万受害者支付共计1.75亿美元的赔偿。）

需要澄清的一点是，在房地产市场繁荣时期累计起来的次级贷款，不管是加州草莓采摘工人的房屋贷款还是巴尔的摩市黑人教民的贷款，并不是数学杀伤性武器。次级贷款是金融工具而不是评估模型，和数学没什么关系。（实际上，贷款经纪人会尽量避免为客户提供任何精确的数字，以免其发现端倪。）

但是，当银行开始把阿尔贝托·拉米雷斯的抵押贷款包装成证券出售时，银行进行此类操作所仰赖的就是有缺陷的数学模型了。此时，资产抵押证券的风险评估模型就是一个数学杀伤性武器。银行知道有些抵押贷款借款人是肯定无法偿还贷款的，但是银行依然坚持了两个错误的假设，以维持它们自己对这个系统的信心。

第一个错误的假设：这些基金和投资机构的顶尖数学家会努力运算大量数据，极其认真地平衡风险。这些债券都是被专家们用最前沿的算法评估过风险的产品。不幸的是，事实不是这样的。和很多数学杀伤性武器一样，数学以造福广大消费者为表象，其实质则是最大化卖家的短期利润。卖家相信他们会在价格下跌前卖掉这些证券。聪明人的确能从中获得盈利，而提供傻瓜资金的“傻瓜们”手里拿的则是总计高达数十亿（甚至万亿）无法兑现的借据。即使是严谨的数学家，其研究的数据也可能是由制造了庞大骗局的人提供的。只有极少的人有能力和必要的信息了解模型背后的算法究竟是怎样的，而大多数业内人士都不愿开诚布公，告诉大家发生了什么。证券风险评级在最初就被设计成一种不透明的、令人望而生畏的复杂系统，这样一来，买家就无从意识到他们手中的债券其风险性的真实水平。

第二个错误的假设：不会有很多贷款人在同一时间违约。这一假设是基于如下理论提出的：贷款违约大部分是随机的、不相关的事件。（事实证明，这一理论毫无根据，很快就被推翻了。）人们因此认为固定抵押贷款可以抵消所有的违约贷款损失。在过去，这一观念的确有其事实依据，而风险模型只假定未来将与过去一致。

为了卖掉这些资产抵押证券，银行需要得到AAA评级。为此，银行需要求助于三家信用评级权威机构。随着市场扩张，对不断增长的抵押贷款债券市场进行证券评级对评级机构来说变成了一笔大买卖，能为其带来丰厚收益。它们热爱由此带来的收益。三家评级机构很清楚，如果它们给出的评级低于AAA，就等于是把一大笔收益拱手让给

了竞争对手。所以，评级机构与银行“狼狈为奸”。评级机构更关注其主要客户（银行）的满意度，而不是模型的精确度。这些风险模型也造成了恶性循环。给次级信贷以AAA评级换取巨额收益，而高收益又造成了人们对这种产品以及整个欺骗造假过程的盲目信任。整个肮脏的商业运作就这样制造出了“互惠互利”的恶性循环，直到崩溃。

在数学杀伤性武器的所有特点当中，将风险模型转化为世界性的破坏力量的是规模化。在之前的房地产泡沫中，被蒙在鼓里的买家最终可能会得到一块沼泽地和大量的虚假契约。而这一次，借助现代电脑运算能力进行的欺诈造成了史无前例的破坏。其他靠资产抵押贷债券发展起来的巨大市场也做出了自己的“贡献”：信用违约掉期市场和综合债务抵押债券（简称CDO）市场。信用违约掉期可被看作一种金融资产的违约保险。这些掉期交易给银行和对冲基金带来了安全感，因为银行和对冲基金理论上可以用掉期交易来平衡风险。然而，一旦持有这些违约保险的机构破产，其造成的连锁反应会使全球经济遭到重创。综合债务抵押债券破坏力更甚。这是一种其价值取决于信用违约掉期以及资产抵押债券表现的债券，其给予了金融工程师们更大的操作权限。

过热（直至崩溃）的资本市场在2007年年底累积了价值三万亿美元的次级贷款，相关市场——包括信用违约掉期市场和综合债务抵押债券市场——中的次级贷款体量则是前者的20倍。没有哪个国家的整体国民经济的体量可以与之匹敌。

矛盾的是，那些创造了这些市场的强大算法，那些分析债务风险并将其包装为债券的算法，在清理混乱局面和计算所有债券的实际价值时变得毫无用处。数学能让谎言和骗局不断繁殖，但不能破译它。这是人类的工作。只有人可以筛选出风险低的抵押贷款，筛掉虚假承诺和一厢情愿的想法，把真金白银投入于购买贷款。这是一个艰苦的过程，因为一方面，与数学杀伤性武器不同，人类无法呈指数级地提

高自己的工作效率，而另一方面，对于整个资本市场来说，完成准确的筛选和投资并不重要。在长期的“排毒”过程中，债务的价值以及债务所依赖的房产价值不断下降。而随着整体经济急转直下，即使是在危机发生时能够负担得起按揭贷款的房主，也可能会突然面临违约风险。

正如我所提到的，德劭集团本来和市场崩溃的震中有一段距离。但随着同行都开始走下坡路，它们疯狂地撤销了那些会直接影响我们客户的交易，这引起了连锁反应。在2008年的下半年，我们公司在各方面都出现了损失。

在接下来的几个月里，灾难终于波及资本市场的核心。大家终于看到了站在算法对立面的人——失去家园的绝望房主以及数百万丢了工作的美国民众。信用卡违约率创历史新高。被数字、电子表格和风险评级遮挡住的人类痛苦变得清晰可见。

2008年9月雷曼兄弟破产以后，公司内部开始讨论整场危机会带来的政治影响。当时，奥巴马有望在11月赢得选举——他会颁布新法规重整产业吗？会对附带权益提高税率吗？我们这些人也许并不会失去房子或者为了继续还贷刷爆信用卡，但大家仍有很多担忧。而唯一的选择就是等待，让说客去做他们的工作，寄希望于继续从前的一切。

到2009年，可以清楚地看到，市场崩溃的教训完全没有给金融界带来新的发展方向，也没有引入新的价值判断标准。大部分说客都成功了，而游戏依然如故：说服傻瓜资金的投入。除了几条新规定之外，一切照常。

整场闹剧让我陡然醒悟。我对数学在其中所扮演的角色极其失望。我被迫面对这个丑恶的事实：从业者只会故意使用公式唬人，而不是用公式来澄清事实。这是我第一次直接面对数学杀伤性武器这个可怖的概念，这让我想要逃避，回到证明题和魔方的世界。

所以，我于2009年离开了对冲基金，深信自己将致力于修正金融领域的数学杀伤性武器。新的法规迫使银行聘请独立专家分析证券风险。因此，我去了为银行提供风险分析服务的公司RiskMetrics集团（简称RMG），其位于华尔街以北的一个街区。我们的产品就是海量数据，每一个数据都旨在预测在下一周、下一年或未来五年，某些证券或大宗商品期货发生亏损的可能性。当几乎每个人都把钱押在了市场中的差不多每一件事物上时，准确的风险分析值得高价购买。

为了计算风险，我们的团队采用了蒙特卡罗模拟法。想象一下，你在赌场里玩了一万次轮盘赌，并且仔细记录下来了每一次的结果。蒙特卡罗模拟法也类似，你通常会从历史市场数据入手，运行数千个测试场景。自2010年或2005年以来的每个交易日，我们所研究的投资组合表现如何？它会在金融危机最黑暗的日子里幸存下来吗？在未来的一两年里，出现致命威胁的可能性有多大？为了给出这些事件的发生概率，数据科学家进行了成千上万次的模拟计算。这种方法有很多不足之处，但它仍是一个评估你的投资组合的风险性的简单方法。

我的工作是担任公司风险管理项目和量化对冲基金（拥有最多最挑剔的风险鉴定专家的机构）经理之间的联络人。我会打电话给对冲基金的管理人员，或者他们会打电话给我，我们会讨论他们对我们给出的数字存在的疑问。但是，在通常情况下，只有当我们犯错误的时候，他们才会联系我。事实上，对冲基金的管理人员一直认为自己是聪明人中最聪明的那群人，并且，因为理解、评估风险是他们存在的根本，他们轻易不会依赖像我们这样的“外部人士”。他们有自己的风险管理团队，他们主要是为了给投资者留下好印象才购买我们的产品的。

我有时也会接到来自大银行客户的咨询。他们渴望恢复在金融危机受损的名誉，希望再次被大众视为负责任的银行，这也就是他们打电话的主要目的。但是，与对冲基金经理不同，他们对我们的分析没

有什么兴趣。他们几乎无视了其投资组合中存在的任何风险。在我与这些客户的整个对话中，我始终难以摆脱一种感觉，那就是警惕风险的人被他们视为“扫兴者”，或者更糟糕，被视为对银行经营底线的一种威胁。即使是在2008年灾难性的金融危机业已发生之后，情况仍未有任何改善。不难理解其中的原因：如果各家银行在如此大的危机中都能幸免于难——因为它们太大而不能倒闭，那么为什么他们现在要担心自己的投资组合存在风险呢?

拒绝承认风险的存在在金融领域已经根深蒂固。华尔街的文化是由交易员定义的，对于风险，他们倾向于尽可能地低估。而这又源于我们对交易员能力的评估方式，即评估他们的“夏普比率”（Sharpe Ratio），该比率是由他们所创造的利润除以其打造的投资组合中的风险计算得来的。夏普比率对交易员的职业生涯、年度奖金、权威性至关重要。如果你将一名交易员视为一组算法，那么，这组算法将不遗余力地致力于优化夏普比率。理想的情况下，随着经验积累，夏普比率会逐步攀升，或者至少不会降至太低。所以，如果一个针对某项信用违约掉期的风险报告针对某个交易员所操作的某类关键资产调高了风险评级，则该交易员的夏普比率就会下降，其年终奖金可能会骤减数十万美元。

我很快就意识到我在做无用功。2011年，到了再次做出改变的时候了，我看到了一个需求像我这样的数学家的巨大市场。我第一次在写简历的时候自称数据科学家，随时准备投身互联网经济。我在一家名为意向媒体（Intent Media）的位于纽约的初创公司找到了一份工作。

我开始建立模型预测各个旅游网站的访问者的行为。模型关注的关键问题是，搞清楚浏览Expedia酒店预订网站的人是单纯地浏览还是真的准备在上面花钱。那些没有消费打算的人的行为对于公司的预期收入几乎毫无价值。对于这部分用户，我们的做法是向其展示比较广

告，比如Travelocity或Orbitz等竞争对手的广告。如果他们点击了广告，我们就会得到几便士——有总比没有好。但是，我们不想将这些广告推送给确实有消费打算的用户。在最糟糕的情况下，我们可能为了一分钱的广告收入而将潜在客户送给了竞争对手，而他们也许会在竞争对手的网站上为预订其在伦敦或东京的酒店房间消费数千美元，这相当于需要成千上万次的广告点击才能弥补回来损失。所以留住这些潜在客户很重要。

我面对的挑战是设计一种算法来区分随机浏览的用户和潜在消费者。有几个明显的信号：他们是否用自己的账号登录了网站？他们以前在网站消费过吗？我也试着搜寻了其他提示，比如一天中的什么时间，一年中的哪些天是消费高峰？一年中的某些时段是用户集中消费的日子。例如，阵亡将士纪念日（5月的最后一个周一）那一天就是一个消费高峰，很多人都会在这一天制订他们的夏季出游计划。我的算法会在这一时段给浏览用户更高的权重，因为他们在这些时段更有可能转化为消费者。

事实证明，对冲基金的金融分析和电子商务的数据统计工作是高度可转换的——最大的区别是，我现在预测的是人们的点击行为，而不再是市场动态。

事实上，我看到了金融产业与大数据产业之间的多种相似之处。这两个行业都吸收了相同的人才，其中大部分来自麻省理工学院、普林斯顿大学或斯坦福大学等精英大学。新晋员工渴望成功，而外界对他们的关注则始终集中于两个关键指标——高考分数和被录取的大学。无论是在金融领域还是在科技领域，他们收到的信息都是，他们将会变得富有，他们将会掌控世界。他们的生产力也表明，他们正走在正确的轨道上，他们将得到实实在在的收益。这些都导向了一个错误的结论：无论采取什么手段，只要赚取了更多的钱就是好的。更高的收入能为其“增值”。不然的话，为什么市场会奖励这种行为呢？

在这两种行业的文化中，财富不再是生活必需品，它直接与个人价值挂钩。一个占尽各种优势的出生于中产阶级社区的年轻人——拥有良好的学前教育背景，在大学入学考试前能得到详尽的升学指导，能轻松获得前往巴黎或上海交流学习的机会——依然自认为，他是依靠自己的能力、辛勤工作和惊人的解决问题的能力进入特权世界的。财富能证明一切。而他的交际圈子则形成了一个相互钦佩、逐步固化的团体。这类人迫切地想让我们相信，这个世界的运作法则是优胜劣汰，但在外人看来，他们只不过是钻了体系的漏洞和运气好罢了。

在这两个行业中，现实世界的复杂性都被搁置一旁，从业者倾向于用数据代替具体的人，致力于把大众转变成更有效的消费者、选民或工人以达成某个自私的目标。当模型成功地以匿名得分的形式出现时，当模型受众对从业者而言仍然只是屏幕上跳动的数字时，这一切都是很容易做到，也很容易被认为是正当且合理的。

当我开始在数据科学领域工作的时候，我便开始写博客了，而且我也越来越多地参与到“占领华尔街”运动之中。我越来越担心技术模型与现实世界中的有血有肉的个体的相互分离的趋势，以及这种分离带来的道德影响。事实上，我在数据、科学领域看到了我在金融领域见到过的同样的模式：一种盲目的安全感导致的不完善模型的广泛应用，模型对于评估效果的自我定义，以及加速恶化的反馈循环。而与此同时，模型的反对者被认为是落后的科技恐慌分子。

我想知道在大数据领域是否存在发生类似于信贷危机的可能性。我看到的不是破产，而是一个越来越糟糕的社会，社会不平等正在加剧。那些复杂的算法被用于确保失败者保持失败者的角色。幸运的少数派将获得对数据经济的更多控制权，掠夺惊人的财富，并说服自己值得拥有这一切。

在大数据领域工作和学习了几年后，我差不多彻底醒悟了，我意识到数学被滥用的程度正在急剧恶化。尽管我几乎每天都在写相关的

博客，但人们被算法操控和恐吓的方式依然令我应接不暇。我们最开始讲的是受苦于教师评估增值模型的华盛顿的中学老师，但故事远并没有就此结束。而我自己因为太过于震惊数学杀伤性武器的破坏性，也已经辞掉工作，开始专心探讨这个问题。

第三章
恶意循环

排名模型的特权与焦虑



如果你在某些城市与朋友一起吃饭，比如旧金山和波特兰，你可能会发现互相分享彼此盘子里的菜是件不可能的事。没有两个人能吃同样的东西。每个人的饮食习惯都不一样。有人是素食主义者，有人推荐原始饮食法，人们恪守不同的饮食标准——至少在一两个月里如此。试想一下，如果其中一种饮食标准，比如说穴居人饮食法，被设定为全美标准，这就意味着3.3亿美国人都要遵循该饮食法的规则吃东西。

这一影响将是巨大的。首先，单一的全美饮食标准将损害农业经济，人们对符合饮食标准的肉类和奶酪的需求将激增，这将大幅抬高该类食品的价格。与此同时，不符合饮食标准的大豆和土豆等作物销路将会变差。食物多样性水平极大降低。遭殃的种豆农民将田地改成养牛养猪场，即使其土地并不适合发展畜牧业或养殖业。而激增的牲

畜将消耗大量的水。不用说，我们大多数人都是绝对不会同意实行单一的饮食标准的。

那么，单一的饮食标准和数学杀伤性武器之间有什么关系？规模化。一种模型算法，不管是饮食方面的还是税法方面的，其在理论上也许是无害的，但是如果将该模型算法推行为全美或者全球标准，其结果就是产生一个扭曲的、极为糟糕的经济体系。高等教育在某种程度上就是这样的情形。

这个故事开始于1983年，美国一家濒临停刊的杂志《美国新闻》决定开展一个规模庞大的项目：评估全美1800所学院和大学，按优秀度为这些学校做一个排名。如果进展顺利的话，这个项目的成果会成为一个有用的工具，可以用于指导数百万的年轻人做好其人生中第一个重大的选择。对于许多人来说，这个选择将决定他们的职业道路、终身的好友圈，通常还包括其终身伴侣。此外，杂志编辑也希望这个大学排名项目能带动杂志销量——没准在推出大学排名的那一周里，《美国新闻》的杂志销量能追上《时代》和《新闻周刊》呢！

但是，要基于什么数据进行大学排名呢？起初，《美国新闻》的工作人员完全依靠他们寄给各大学的校长的调查问卷所得到的反馈结果进行评分。结果，斯坦福大学位居全美综合性大学之首，阿默斯特学院则是排名第一的文科学院。排名结果虽然很受读者欢迎，但也令很多大学的校领导非常愤怒。杂志社收到了排山倒海般的投诉，内容都是控诉排名结果有失公正的。许多大学的校长、在校学生和已毕业的校友坚持认为自己的学校应该获得更高的排名，杂志社应该再去仔细研究一下有关的数据。

接下来的几年，《美国新闻》的编辑一直在思考他们具体可以测量什么数据。许多模型诞生了，但其中大量的评估因素仅仅来自直觉。模型确立的过程并不严谨，统计分析也缺少根据。在大学排名这个案例中，模型建立的依据仅仅是人们凭空想象什么是对教育而言最

重要的因素，然后，这些人便去寻找可以测量的相关变量，最后随意地在公式中赋予每个变量一定的权重，这样模型就完成了。

在大部分领域中，模型确立的过程通常还是比较严谨的。比如说，农业学科的研究者会比较投入（土壤、阳光和化肥）和产出（收获后，具有特定特征的农作物的产量）。然后，他们就可以按照他们的目标，比如一定的成本、口感或者营养价值等进行下一步的试验和优化。但这并不是说农学家就不可能建立数学杀伤性武器。他们也曾经建立过数学杀伤性武器（尤其是当他们没有考虑到杀虫剂长期而广泛的影响时）。但是，因为他们的模型在大多数情况下关注的是一个明确的结果，所以我们仍然认为这是一个相对理想的科学实验过程。

但是，《美国新闻》的编辑所做的是“教育优秀度”排名，这是比粮食成本或者每个麦粒的蛋白质含量更加抽象、模糊的价值。这些编辑没有直接的方法来量化4年的大学学习过程是如何影响某一个学生的，更不用说数千万个学生了。他们不可能测量一个学生在4年的大学生活中的学习、幸福、信心、友谊等全部方面。美国前总统林登·约翰逊对高等教育的定位是：“高等教育是深化自我实现、扩大个人生产力和增加个人回报的途径”，但不管是其中的哪一条都不适合放在大学排名模型中。

《美国新闻》的编辑只是挑选了一些和评估目标看似相关的变量。他们研究了高中生的SAT（学业能力倾向测验）成绩、学校的学生教师比和录取率。他们统计了顺利进入大二的新生占总数的百分比和顺利毕业的学生占总数的百分比。他们计算仍在世的已毕业校友为母校捐款的人数占总数的百分比，依据是他们给母校捐款很可能表明他们喜欢母校的教育。排名结果中占3/4权重的分数都来自一种算法——一种被公式化了的主观观点，这种算法就包含以上那些变量；另外占1/4权重的分数则来自全美各地的大学校长的主观评价。

《美国新闻》第一次依据数据确定的大学排名于1988年出炉，从表面上看，排名结果没有太大问题。但是，当这一排名发展成全美标准时，恶性循环出现了。关键问题就是，排名会自行巩固。如果一所大学在《美国新闻》所发布的排名中名次靠后，它的声誉就会下降，生源情况就会恶化。优秀的学生会避开这所大学，优秀的教授也一样。已毕业的校友将减少捐款。由此一来，这所学校的排名就会继续下跌。简单来说，排名决定了大学的命运。

以前，大学的校领导有各种方法宣扬学校教育的成功，许多都是靠传闻逸事。例如，某些教授得到了众多学生的一致好评；一些学生在毕业后走上了杰出的职业之路，成为外交官或者成功的企业家；还有一些学生出版了一流的小说。这些正面事迹经由口口相传广为人知，学校的声誉也由此提升。但是，麦卡利斯特学院就比里德学院好吗？或者艾奥瓦大学就比伊利诺伊大学好吗？这很难说。不同的大学就像不同类型的音乐或者不同的饮食习惯，对于某所大学的评价众说纷纭，好坏两方面都可以列出充分的理由。而现在，大学的整体声誉生态系统被一组数字蒙上了阴影。

如果你站在大学校长的角度思考这件事情，你会发现大学排名其实是很糟糕的。毫无疑问，绝大多数校长珍惜自己的大学经历，因为从某种程度上来说，正是大学经历激励他们攀登学术阶梯，成为一所大学的校长。但是现在，这些正处在事业高峰期的校长需要投入巨大的精力提高与学校教育优秀度有关的15个考核项的分数，而这15个考核项是由一个二流杂志社的一组编辑定义的。他们就好像又回到了学生时代，每天都在祈求老师给高分。实际上，他们正是掉进了死板模型，即数学杀伤性武器的陷阱之中。

如果《美国新闻》发表的大学排名只在小范围内流行的话，倒也不会造成什么麻烦。但是，这个排名的影响力发展迅速，其很快成为一个全美标准。教育界一下子紧张起来，迅速给大学校长和学生都设

定了严格的任务清单。《美国新闻》的大学排名模型规模巨大，造成了大范围的损害，导致了几乎是无尽的恶性循环。尽管这种模型不像许多其他的模型那样不透明，但它确实是一个数学杀伤性武器。

一些大学的校领导想尽一切办法提高排名。贝勒大学设立奖金激励大一新生再次参加SAT考试，希望再考一次能提高他们的成绩以及贝勒大学的排名。有些名校，包括宾夕法尼亚州的巴克内尔大学和加利福尼亚州的麦肯纳学院，则给《美国新闻》反馈了假数据，夸大了其学校新生的入学分数。2011年，位于纽约的爱纳学院承认其学校教师几乎捏造了所有的数据：考试成绩、录取率和毕业率、新生保留率、师生比和校友捐赠额。但谎言起效了，至少在一段时间之内。据《美国新闻》估算，假数据将爱纳学院从东北地区大学排名第50名提升至第30名。

更多的校领导则试图寻找一种更常规的方式来提高他们的学校排名。他们没有作弊，而是努力提升学校在影响最终分数的每一个变量上的表现。他们可能会认为这是效率最高的资源利用方式。毕竟，只要他们努力去迎合《美国新闻》的模型，得到更高的排名，他们就能筹集到更多的资金，吸引来更优秀的学生和教授，然后进一步提高排名。除此之外，他们还有别的选择吗?

罗伯特·莫尔斯从1976年起就在《美国新闻》杂志社工作了，他是这个大学排名项目的组织者，他在采访中称进行大学排名有利于推动大学制定更有意义的目标。如果他们能因此致力于提高毕业率或者把学生分成更小的班级上课以提高教学效果，那就说明排名是件好事情。他承认杂志社拿不到与大学教育优秀度最相关的数据，即每个学校学生的学习内容。但是，基于替代变量建立的《美国新闻》大学排名模型也足够反映问题了。

但是，当你基于替代变量建立模型时，钻模型的漏洞会变得容易很多。这是因为替代变量比起其所代表的复杂事实更容易操控。举个

例子，假设有一个网站要聘用一个社交媒体专家。许多人前来应聘这个工作，他们发送来自己以往运营的各种营销活动的相关信息。但是，这些信息需要大量的时间去证实，而据此评价他们的工作表现也并不简单。所以，人事经理决定选定一个变量——重点考虑推特粉丝数排名靠前的应聘者。推特粉丝数是社交媒体参与度的标志之一，没错吧?

这是一个很合理的替代变量。但是想象一下，当推特粉丝数是获得该公司职位的关键这一消息遭到泄露时会发生什么？应聘者很快就会无所不用其极地增加推特粉丝。有人会花费19.95美元直接“购买”大量由机器操控的粉丝。因为人们钻了招聘系统的漏洞，替代变量失去了效力。

在《美国新闻》大学排名事件中，从高中毕业生到大学校友再到公司的人力资源部，人们很快接受了该排名是大学教育质量的一个体现。因此，各个大学只能选择配合，他们不得不努力提高排名所涉及的每一个考核项的分数。其实，许多学校最焦虑的是那不能控制的占排名结果1/4权重的因素，即声誉分数，来自各个大学、学院的校领导给出的问卷调查反馈。

这占据1/4权重的部分，就和所有的主观观点的集合一样，必然包含着过时的偏见和无知。位居排名前列的知名学府往往也会得到一致好评，因为人们熟知这些学校。而对于还未被人熟知、渴望占据一席之地的学校而言，提升排名则变得更难了。

2008年，沃思堡市的得克萨斯基督教大学（TCU）排名猛降。三年前，该校的名次是97，之后三年名次递降为105、108和113。该校的校友和支持者为此感到很愤怒，校长维克多·博西尼也因此陷入尴尬境地。“这整件事让我感到很沮丧。”博西尼对学校的校园新闻网表示。他坚称得克萨斯基督教大学在每个指标上的表现都在进步，“我们的新生保留率在提高，我们的筹款等所有方面都在改善”。

博西尼的申辩有两个问题。首先，《美国新闻》排名模型并不是对各个大学进行孤立的判断。即使是各指标分数均有所提升的学校在排名中也会落后于其他分数提升得更快的学校。用学术术语来说，《美国新闻》的评估模型是一种分布模型。这导致了一场学校间的军备竞赛。

另一个问题是，得克萨斯基督教大学无法控制占1/4权重的声誉分数。招生主任雷蒙德·布朗指出，声誉是模型中权重最大的变量，“这很荒谬，因为它完全是主观的”。新生招生主管威斯·瓦戈纳则指出，为了提高声誉分数，各大学都在纷纷为自己打广告。瓦戈纳说：“我甚至收到过来自其他大学的邮件，其内容是试图说服我他们是一所好学校。”

尽管如此，得克萨斯基督教大学仍然决定着手提升那可控的占3/4权重的分数。毕竟，如果各项指标的分数均有所提升，大学声誉最终也会随之提升。随着时间的推移，同行们会注意到它的进步并给予其更高的评价。关键是要走上正确的轨道。

得克萨斯基督教大学发起了一个2.5亿美元的筹款活动。到2009年，学校已募集到4.34亿美元，远远超过目标额度。由于筹款额是排名的指标之一，仅此一项成绩就提升了得克萨斯基督教大学的排名。得克萨斯基督教大学花费了其中的大部分资金用于校园设施改善，其中1亿美元用于兴建中央商场和学生活动中心，努力让得克萨斯基督教大学的校园看上去更具吸引力。这些做法本身没有什么不对，但其初衷是迎合《美国新闻》的排名模型。毕竟，申请的学生越多，学校就越有选择的余地。

也许更重要的是，得克萨斯基督教大学兴建了一个其时最高水准的体育训练场馆，并将大量的资源投入到其足球项目之中。在接下来的几年里，得克萨斯基督教大学的角蛙足球队成为国家强队。2010

年，他们在玫瑰杯足球赛中打败了老牌强队威斯康星队，取得了全美总冠军。

这次胜利为得克萨斯基督教大学带来了所谓的“弗洛特尔效应”（the Flutie effect）。1984年，在一场极为精彩的大学橄榄球比赛上，波士顿大学队的四分卫道格·弗洛特尔在最后一秒完成了一个扭转败局的超长距传球，打败了迈阿密大学队。弗洛特尔由此成为一个传奇。这场比赛结束后的两年内，波士顿大学的大学申请率上涨了30%。乔治城大学也曾拥有带来过同样的宣传效果的传奇。该校由帕特里克·尤因带领的篮球队三次打进全美锦标赛。看来，赢得体育比赛是吸引学生申请某所大学的关键因素。当大批体校的高三学生在电视上观看大学体育比赛时，球队实力强劲的学校对他们形成了极大的吸引力。这些学生会为自己是该校的学生、身着写着该校校名的队服而感到骄傲。这些大学接到的入学申请因此暴涨。随着更多的学生申请入学，招生处就可以提高入学门槛，以提高大学新生的SAT平均分，而这有助于提高大学排名。另外，学校拒绝的申请学生越多，其录取率就越低，对排名就越有利。

得克萨斯基督教大学的策略奏效了。到2013年，该大学已成为得克萨斯州学生选择度排名第二的大学，排在第一的是著名的休斯敦莱斯大学。这一年，得克萨斯基督教大学的新生高考和入学考试平均成绩均达到史上最高水平，其在全美的排名也因此大幅上升。2015年，该校全美排名76，也就是说，仅用了7年时间，该校就上升了37个名次。

尽管我对《美国新闻》的大学排名模型有异议，且该模型确实是个数学杀伤性武器，但是需要指出的是，得克萨斯基督教大学的排名跃升对其自身发展是有切实帮助的。毕竟，《美国新闻》的排名模型中的大多数变量在某种程度上的确反映了一所学校的总体质量，就像确实有不少节食者因遵守穴居人饮食法而节食成功了一样。问题不是

出在《美国新闻》排名模型本身上，而是出在该模型的规模上。该模型迫使每个人、每个学校都认准同一个目标，这导致了激烈竞争，以及很多意料之外的有害后果。

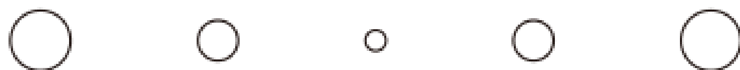
比如说，在这个大学排名项目启动之前，准大学生在知道自己申请到了一所入学标准较低的所谓保底大学之后，就不会太过担心。如果学生没有进入他们的首选大学，包括略高于其实际水平的冲刺大学和与其实际水平相符的目标大学，则他们在保底大学也可以接受较好的教育，并且有机会在一两年后转到一所他们的首选学校。

但现在，保底学校的概念已经基本消失了，这主要归咎于《美国新闻》的排名模型。正如我们在得克萨斯基督教大学的案例中看到的那样，大学排名有助于提高学校的学生选择度。如果一所大学的招生办收到的入学申请泛滥成灾，这就是学校走对了路的信号，证明该所大学声誉极佳。同样，如果一所大学拒绝了绝大部分的入学申请，它最终的新生SAT平均水平必然会更高。和许多替代变量一样，这一指标似乎非常合理，遵循了市场的变动规律。

但是，这个市场是可以操控的。例如，一所传统的保底学校通过查看历史数据可能会发现，只有一小部分优秀的申请者被录取了，而余下的大多数都进入了他们的目标学校或冲刺学校。因此，为了提高自己在学生选择度方面的评分，这所保底学校现在更倾向于拒绝那些——根据其历史数据推导出的——最有可能被更好的大学录取而放弃该校的优秀考生。整个筛选程序可以说是本末倒置的。无论招生办公室的数据专家怎样卖力工作，这所大学最终无疑仍会丢失一定量的本想选择该校的优秀学生。同时，让学生感到沮丧的是，所谓的保底学校也不再是一个安全牌。

这个复杂的过程对提升教育效果没有任何帮助。大学遭受了重大损失，顶尖学生大量流失。事实上，以前的保底学校现在可能不得不

设置助学金才能吸引那些优秀的学生。而这对那些最需要助学金的学生来说，则意味着其学费负担更重了。



现在，我们终于发现《美国新闻》大学排名模型最大的缺陷是什么了。我们不能说《美国新闻》的编辑为评判“教育优秀度”选择的替代变量是无效的，但他们犯下的更大的错误来自他们没有纳入考虑的变量：学杂费、学生助学金。这些变量被该排名模型遗漏了。

这引出了我们在本书将会频繁讨论的一个关键问题：建模者的目标是什么？在大学排名这个案例里，你需要站在1988年《美国新闻》的编辑们的角度来考虑。当他们在建立第一个统计模型的时候，他们怎么知道这一模型是否有效？首先，如果模型能反映一些已有定论的大学排名，这就表明其有一定的可信度。比如，如果哈佛大学、斯坦福大学、普林斯顿大学和耶鲁大学在大学排名模型中位居前列，这就在一定程度上证实了《美国新闻》编辑设计出的大学排名模型是有效的。而要建立这样一个模型，他们只需要去研究那些一流高校，思考这些大学的特殊之处是什么就可以了。优秀大学的共同点是什么？这些学校与其隔壁镇的保底学校差距何在？他们发现：优秀大学的新生SAT成绩都很高，而且绝大部分都能顺利毕业；已毕业的校友都很有钱，会不断给学校捐款；等等。就这样，《美国新闻》的大学排名项目组通过分析名牌大学的优势，建立了一个测量教育优秀度的评估指标体系。

现在，如果该项目组将教育成本纳入算法，则其模型输出也许会发生奇怪的变化——学费便宜的大学很可能因此闯入优秀大学之列，而这一结果将遭到广泛的质疑。由于公众可能会把《美国新闻》最终公布的大学排名看得特别重要，因此采取保守、常规的算法，保证一

流大学位于排名输出结果的前列，是一种更安全的做法。当然，高成本也许正是优秀的代价，这也不是没道理。

《美国新闻》的排名模型把成本排除在算法外，这就好像是给大学校长们递了一本镀金支票簿。后者要遵循的唯一指令，就是最大限度地提高15个考核指标的评分，而降低成本则不在其列。事实上，提高学费反而能让他们有更多的资源用于提升考核项目的表现。

从此，学费一路飙升。从1985~2013年，高等教育的学费上涨了5倍以上，差不多是通货膨胀率的4倍。为了吸引顶尖的学生，各大学都像得克萨斯基督教大学一样，纷纷开始大力投入校园基础建设，建造有玻璃墙的学生中心、豪华的宿舍，以及带攀岩墙和漩涡浴缸的健身房等。从表面来看，这对学生来说是好事，这些设施可以丰富他们的大学体验——前提是他们不需要以助学贷款的形式承担这些费用，偿还助学贷款的压力可能会跟随学生几十年的时间。不过我们不能把一切都归咎于《美国新闻》的大学排名。我们整个社会不仅认同了大学教育是必不可少的这一观念，而且欣然接受了排名靠前的学校的文凭能帮助学生快速进入特权阶层这一事实。《美国新闻》的排名模型以由此而生的恐惧和焦虑为养分成长为一个庞然大物。排名模型有力地刺激了各方在教育上的不断投资，而飙升的学费则被忽视了。

出于提升排名名次的需要，各个大学就像管理投资组合一样管理着自己的学生。这在大数据领域里很常见，小到广告业大到政治领域都是如此。在校领导看来，每一个准大学生都代表着一组资产和一两项债务。比如，一名高中生在体育赛事上的优秀表现就被视为一种资产，但同时她的成绩可能处于中下游水平，后者就是她背负的债务。她可能还需要申请助学金，这又是一项债务。为了平衡投资组合，他们应该发掘其他能自费上学并且成绩优秀的考生。但是那些理想考生即使被录取了也可能会选择去其他更好的学校。这也是一个必须要量

化的风险。鉴于整个评估体系非常复杂，为了“优化招生”，教育咨询产业兴起了。

教育咨询公司诺埃尔–莱维茨（Noel-Levitz）开发了一个被称为“预告+”（ForecastPlus）的预测性分析软件包。该软件包允许招生老师根据地理位置、性别、种族、研究领域、学术地位及“任何其他特征”对准大学生的情况进行评估。另一个名叫“定位学生”（RightStudent）的咨询机构则致力于收集、买卖相关数据以帮助大学客户找到最适合录取的学生人选，包括可以支付全额学费的学生，以及可能有资格获得校外奖学金的学生。就这个意义而言，学习障碍对于大学录取可能反而是个优势。

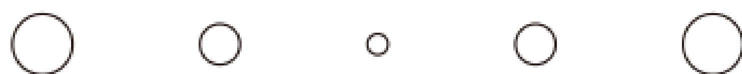
所有这些都发生在这个以《美国新闻》大学排名为中心的巨大的生态系统里，排名模型实际上充当了系统内部最高法的角色。如果《美国新闻》的编辑重新安排模型中部分替代变量的权重，比如降低考试成绩的权重，或者增加毕业率的权重，则整个教育生态系统就要重新适应新的法则。这一改变将波及咨询公司、高中的升学指导部门，以及所有的学生。

排名自然而然地成为一个不断自我巩固、自我发展的特权。2010年，《美国新闻》杂志停刊了。但是整个排名产业未受丝毫影响，且继续发展壮大，排名延伸到了医学院、牙科学校、文学和工学研究生院，甚至高中。

随着排名产业的发展，钻模型漏洞的手段也越发丰富。2014年的《美国新闻》全球大学排名中，沙特阿拉伯的阿卜杜勒阿齐兹国王大学（KAU）的数学系排名第7，仅次于哈佛。然而该校的数学系仅成立了两年，没人知道它是如何一下子跃升至全球前10，甚至超过了剑桥大学和麻省理工学院的数学系的。

乍一看，这似乎是一种积极的转变。也许麻省理工和剑桥只是靠它们的好名声取得好名次的，而KAU则是凭实力上位的。如果只参考学校声誉来排名，这样的排名逆转需要几十年的时间才会发生，而大数据模型缩短了变化呈现的时间。

但是，算法本身也有能被钻空子的漏洞。伯克利大学的电脑生物专家利奥·帕赫特研究了这个问题。他发现，KAU和论文引用次数极高的很多数学家进行了接触，并以7.2万美元的年薪聘请他们担当该校的客座教授。根据帕赫特找到的招募信，该合作协议规定这些数学家每年必须在沙特阿拉伯工作三周。大学将承担他们的商务舱机票，安排他们入住五星级宾馆。可以想见，他们在沙特阿拉伯的工作为学校增加了价值。但更关键的是，该大学还要求这些数学家将他们记录在汤森路透学术引用网站上的通信地址改为KAU，而这正是《美国新闻》排名模型中的一项关键参考因素。这意味着，KAU可以声明他们的众多新任客座教授的学术论文和专著都是他们的成果。由于论文引用次数是排名模型算法里的一个重要参考数据，KAU因此排名飙升。



中国湖北省钟祥市的学生曾以高考成绩极佳享有盛名，全国一流大学都有他们的身影。但他们的表现好过头了，有关部门开始怀疑当地存在高考作弊的现象。据英国《邮报》的一个报道，2012年，当地省级机关发现在一次模拟考试中，有99份答案一模一样的试卷。

第二年，钟祥市的学生被要求在进入考场前通过金属探测仪，并交出手机。一些学生被发现携带了伪装成橡皮擦的小型信号传递器。一进考场，学生们发现自己面对的是来自不同学区的54个考场侦查员。一些侦查员在考场附近的一个旅馆里，发现一群人正准备通过信号传递器给考生传输试卷答案。

这次作弊打击行动引发了各方的激烈反应。约2000名愤怒的学生家长聚集在学校门口抗议。他们喊道："我们要公平，不让作弊就没法公平。"

这听起来像个笑话，但他们绝对是认真的。因为赌注太大了，摆在学生们面前的只有两条路，进入名校读大学然后找到一份体面的工作，或者一辈子待在这个相对落后的小城市里。而且不管事实如何，他们认为其他学生也都在作弊，所以不让他们作弊就是不公平。在一个作弊成风的制度下，遵守规定反倒成了不利条件。兰斯·阿姆斯特朗和他的队友靠着兴奋剂连续7年打败了环法自行车比赛中的其他选手，那些同样遵守规定的选手想必对此感同身受。

在这种情况下，唯一的胜出方式就是牢牢掌握某个优势，同时确保其他人的优势不比你的更大。集体性的考试作弊不仅在中国有，在美国也有。不管是高中招生办还是家长、学生，所有人都疯狂地想要钻大学排名模型的空子。

排名模型生产的恶性循环及其引起的广泛焦虑，也导致了整个升学辅导教育产业的蓬勃发展。一个叫"名校录取"（Top Tier Admissions）的教育公司推出了一个为期4天的"大学申请训练营"培训项目，收费高达1.6万美元（不包括住宿和饮食）。在这期间，这些高二学生将学习如何写申请书，学习如何"拿下"面试，创建"活动列表"总结自己得过的所有奖项和参加过的所有体育运动、社团活动及社区志愿服务，因为这些都是大学招生办关注的要素。

1.6万美元听起来似乎很贵，但就像中国钟祥市的家长一样，很多美国家庭愿意付出这样的代价，因为他们认为孩子未来的成功和成就取决于是否能被名牌大学录取。

教育公司的专业人士了解每个学校的招生模型，所以他们知道怎样让一个准大学生被纳入其目标学校的"投资组合"之中。一位加州的

企业家在教育产业把市场分析法发挥到了极致。他叫马振翼，是美国星腾科国际教育集团的创始人。他用自己开发的模型评估准大学生，计算他们被目标院校录取的可能性。他对《彭博商业周刊》的记者表示，假设一个美国高中生的平均学分绩点（GPA）为3.8，SAT成绩为2000分，课外活动时间为800小时，那么他被纽约大学录取的概率为20.4%，被南加州大学录取的概率为28.1%。然后，星腾科将提供一份有担保的建议组合。如果这个学生接受了咨询公司的建议辅导并最终成功被纽约大学录取，则该学生就需向咨询公司支付25931美元，如果他最终成功被南加州大学录取，则需要支付18826美元。如果他的申请被两个学校都拒绝了，那么咨询公司将不收取任何费用。

每所大学的招生模型都全部或者至少一部分来源于《美国新闻》的大学排名模型，这些模型中的每一个都是一个小型的数学杀伤性武器。正是这些招生模型让学生和家长身陷焦虑，花掉大把的钱。而且这些招生模型都是不透明的，大多数的申请学生（或者叫受害学生）都被蒙在鼓里。这就为像马振翼这样的专业咨询人士创造了巨大的商机，通过培养其在各个大学的人脉以获取第一手信息或者逆向推导各个学校的招生模型算法，他们破解了绝大部分学校的招生模型。

当然，主要受害者仍然是美国的大多数，即穷人和中产阶层，他们没有那么多钱可以花在课程和咨询公司上。他们错失了珍贵的内部信息。结果是，教育体系偏向于特权阶层，偏离于穷人和中产阶层，淘汰后一类家庭出身的绝大多数学生，将他们推向贫穷之路，进一步加剧了社会阶层固化。

但是，即使是那些想尽办法进入了名牌大学的学生也并不是赢家。大学招生制度只对少数人而言是有利可图的，且根本没有任何教育价值，只不过是以某种新奇的方式将一群18岁的孩子重新排序分类。在备考阶段，掌握更多的篮球技巧或者在专业辅导人员的帮助下写出符合目标大学标准的申请书并不能让他们掌握真正有意义的技

能。更不用说很多人都是靠蒙混过关的。所有这些学生，不管是来自富人阶级还是来自工人阶级，都被培训成要去适应一台巨大的机器，一个被大规模投入使用的数学杀伤性武器。严酷的考验结束后，很多人得到的只是累累负债，他们是这场极为险恶的“军备竞赛”中的无名小卒，只能受人摆布。

那么，有什么解决办法吗？在奥巴马总统的第二任期内，他提出了一个新的大学排名模型，比《美国新闻》的大学排名模型更符合占全美大多数的中产阶级的利益。他的次级目标是削弱营利性大学（这是一个吸钱祸害，我们将在下一章讨论）的影响力。奥巴马的想法是将大学排名系统与一组不同的指标联系起来，这些指标包括负担能力、贫困学生和少数族裔学生所占比例以及学生毕业后的就业情况。和《美国新闻》的排名模型一样，该模型也会考虑毕业率。如果某所大学在这些指标上的表现低于最低标准，其就会被踢出每年价值1.8亿美元的联邦助学贷款市场（营利性大学一直身在其中）。

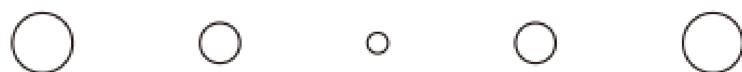
奥巴马的建模目标听起来很有价值，但是每一个排名模型都有漏洞可钻。而一旦被钻了空子，模型就会产生新一轮的恶性循环以及大量意料之外的有害后果。

举例来说，提高毕业率很简单，只需降低毕业要求就可以了。许多学生无法通过数学、科学专业课、外语这几门课的考试，那么放宽这方面的要求，这样更多的学生就能毕业了。但是，如果我们的教育体系的目标是培养更多的科学家和技术人员，那么这种做法岂不是很讽刺？提高毕业生的收入水平也很容易办到。所有大学要做的就是减少文科专业，撤掉教育系和社会服务系，因为教师和社会工作者挣的钱没有工程师、化学家和计算机科学家多——虽然前者对社会而言必不可少。

降低学校成本也不是太难。一个已经广泛流行的方法就是降低终身教授在学校教职工中所占的比例，在他们退休后聘请成本较低的讲

师或者兼职教授。对某些大学的某些院系，这可能是有效的。但是，这是要付出代价的。终身教授会带领研究生做具有重要意义的研究，以自己的工作表现为其所在系设定标准，而忙碌的兼职教授可能为了交房租在三所大学教五门课程，几乎不可能有时间或者精力为学生提供更好的教育。还有一个办法是撤掉一些不必要的行政职务，但这种做法似乎太罕见了。

“在毕业后9个月内就业的毕业生”数量也可以造假。《纽约时报》2011年的一份报告针对法学院做了一项调查，该报告评估了各个大学的法学院对毕业生的就业安置能力。调查显示，假如一位肩负15万美元学生贷款的法学院毕业生只找到了一个在咖啡店打工的工作，那么一些无良的法学院会把这个学生也计入就业人数。另一些学校更过分，在9个月的期限即将截止之时，对于那些还没找到工作的毕业生，学校就雇用他们在学校做小时工，并将其计为就业。还有学校向毕业不久的校友发出调查，并将所有没有回复的人都归为“就业”。



或许，奥巴马政府没能拿出一个经过重新调整的排名系统也是好事。大学校长强烈抵制新的排名系统。毕竟，他们多年来一直是照着符合《美国新闻》的排名模型的方向努力的。奥巴马提出的新排名模型涉及毕业率、班级人数、毕业生就业安置情况和收入水平等其他变量，若严格按照指标标准评估各大学，其得出的评分将严重损害众多大学的排名和声誉。毫无疑问，他们相当熟悉这类模型的缺陷和其可能产生的恶性循环。

所以，政府最终做出了让步。也许这一妥协后的结果比推行新模型更好。教育部没有将大学重新排名，而是把大量的调查数据公布在网站上。这样一来，学生就可以自行查询自己关心的指标，包括班级人数、毕业率以及应届毕业生的平均负债额等。他们无须再去了解任

何统计法或者变量的权重。就像一个旅游网站一样，每一个人可以自行制定个人的模型。想想看：透明，用户控制，个人化——完全是数学杀伤性武器的对立面。

第四章
数据经济

掠夺式广告的赢家



在我担任初创广告公司意向媒体的数据科学家期间，有一天，一位著名的风险投资人来公司访问。他似乎正在考虑投资我们公司，因此，公司急切地想要表现出自己最好的一面，所有员工都被要求去听他讲话。

他向我们描述了定向广告的辉煌前景。人们在互联网上贡献了海量数据，广告商因此可以事无巨细地了解他们，从而用针对每个用户量身定做的营销信息锁定客户，且营销信息会在合适的时间、合适的位置精准投放。举个例子，比萨店不仅能知道你就住在店面附近，而且能知道你这会儿正想要吃深盘意大利香肠双层奶酪比萨，就像你上周在看电视直播的达拉斯牛仔队比赛的中场休息时间做的那样。比萨店的系统会发现，行为模式和你相似的人更有可能在中场休息的20分钟里点击领取网站上的打折优惠券。

我觉得，这位风险投资人的论点最薄弱的部分是理由。他认为，在不久的将来，铺天盖地的个性化广告将会因为有效且及时而广受顾客欢迎。因此，他们会想要更多。他发现，大多数人排斥广告是因为大部分广告和他们关系不大。而在未来，情况将会发生变化。在他的假定里，人们会喜欢商家为他们量身定做的广告，比如巴哈马群岛的乡间小屋，手工制作的初榨橄榄油，或者分时共享的私人飞机。他还开玩笑说，他再也不用看到菲尼克斯大学的广告了，这所大学是一家营利性的教育工厂，其招收的大部分学生都来自挣扎谋生（和更好骗）的底层阶级。

我觉得他提及菲尼克斯大学的广告这事很奇怪。说来也怪，他看到了该所大学的广告，而我却没有，或者是我没有注意到。我对营利性大学可以说很了解，它们运作的是高达数百万美元的交易。这些所谓的“文凭工厂”通常是由政府提供贷款并予以担保的，但其颁发的文凭在职场中并没有多少价值。对于很多职业而言，这种文凭并不比高中学历更有价值。

一边是《美国新闻》的大学排名模型刺激富人和中产阶级学生及其家庭为申请大学投入巨资，而另一边，营利性大学则把矛头指向另一类更脆弱的群体。互联网为其提供了一个完美的工具。因此，该行业的迅猛发展与互联网这个永远在线的大众交流平台的发明和流行相同步也就不足为奇了。尽管其仅在谷歌广告投放这一项上就花了5000多万美元，但菲尼克斯大学的目标是穷人，而它放出的“诱饵”是一个突破固有阶层上升一步的机会。这个“诱饵”还暗含着对其目标的潜在批评，即认为这些底层阶级没有付出足够的努力去改善自己的生活。这个“诱饵”起效了，穷学生们上钩了。从2004年到2014年，营利性大学的招生人数增长了3倍，目前，营利性大学的招生人数占全美大学招生总人数的11%。

营利性大学的市场营销，与互联网早期作为一种推行平等和民主的巨大力量的承诺相去甚远。如果说在早期的网络时代，“没人知道你是一只狗”，那么今天的情况正好相反。互联网基于我们的网上行为所透露出来的内在偏好和选择模式，把我们放在数百种模型中进行排名、分类以及评分。这为合法的广告营销奠定了强大的基础，但同时也助长了掠夺式广告的兴起，后者指精确找出有迫切需求的群体，并向他们兜售虚假承诺的广告。掠夺式广告以寻找不平等并大肆利用不平等为己任，其结果是进一步巩固了现有的社会分层。整个社会最大的分水岭就出现在这个系统的赢家（比如风险投资人）和赢家所建立的模型的掠夺目标之间。

在人们既有迫切需求又对具体信息很无知的任何地方，你都能看到掠夺式广告。如果人们对他们的性生活感到焦虑，那么掠夺式广告商就会向他们兜售伟哥或西力士，甚至阴茎延长术；如果人们缺钱，那就向他们兜售高息发薪日贷款；如果有人发现自己的电脑运转缓慢，那么这台电脑很可能中了一个由掠夺式广告携带的病毒，其目的是制造购买修理服务的需求。我们看到，营利性大学的急速发展正是受助于掠夺式广告的兴起。

掠夺式广告其实就是一种数学杀伤性武器。它们大规模聚焦于社会中最绝望的那群人。在教育方面，它们承诺通向成功的道路，但这条路的终点往往是贫穷的深渊，而与此同时，它们则在计算如何从每一个准大学生身上最大化地榨取利润。掠夺式广告造成了一个巨大的恶性循环，导致它们的客户债台高筑。而且，目标群体对自己是怎样被欺骗的知之甚少，因为这些广告营销的投放机制是不透明的。他们所看到的只是广告在电脑上突然出现，然后电话就打来了。受害者几乎无从了解他们是如何被选中的，或者广告商是如何能够对他们有那么详尽的了解的。

以科林斯大学为例。直到3年前，这所学校仍是营利性大学的巨头。该大学各个分校总共有80000多名学生，其中绝大多数学生都接受了政府贷款资助。2013年，加州司法部部长起诉该营利性大学谎报就业率，收取过高学费，在广告中使用造假的军用印章欺骗弱势群体。该起诉书指出，该大学的一个分校针对律助专业的函授本科录取学生收取高达68800美元的学费。（这类函授学位的学费在全美大部分传统大学为不到1万美元。）

此外，根据起诉书，科林斯大学的营销目标群体是“孤立”“缺乏耐心”“低自尊”的人，这些人在“生活中很少被他人关心”，他们往往感觉自己被“困住”，“没有对未来的期许和人生规划”。起诉书称科林斯大学的做法是一种“非法的、不公平的欺诈”行为。2014年，由于出现了越来越多的有关该校的恶劣行径的报道，奥巴马政府撤销了科林斯大学申请联邦助学贷款的资格，这切中了科林斯大学的命脉。2015年年中，科林斯大学出售了其大部分的校园土地，并宣布破产。

但营利性教育产业仍在继续发展。职业培训机构翡特罗特学院（Vatterott College）就是一个性质极其恶劣的案例。2012年，参议院委员会在一份关于营利性大学的报告中描述了翡特罗特学院的招生手册，称其“内容极为残忍、恶毒”。手册指导招生人员锁定具有如下特征的群体：以社会救济为生的带着孩子的单亲妈妈，怀孕的妇女，刚刚遭遇离婚，低自尊，从事低收入的工作，最近经历过身边人死亡的事件，遭受过身体/精神虐待，近期有入狱经历，有药物成瘾治疗经历，从事没前途的工作——用一句话概括，就是没有未来的人。

为什么营利性大学要锁定这些人呢？因为对它们而言，脆弱的价值好比黄金，一直都是如此。老式西部电影中往往有一个巡回庸医的角色。这个人会驾着马车进入城镇，马车里装满叮当响的瓶瓶罐罐。当他与一位老婆婆——他的潜在客户坐在一起时，他会依据种种迹象找出她的弱点。当她微笑时，她会捂住嘴，这表明她对她的坏牙很敏

感。她不安地转动着手上的旧结婚戒指，从她那肿胀的关节来看，戒指会一直卡在那儿直到她死的那一天。这说明什么？她有关节炎。因此，当他向她推销产品时，他就可以向她特别强调她牙齿的丑陋和她那酸痛的手，并向她承诺可以让她恢复美丽的微笑，并消除她关节的疼痛。在得到了这些信息之后，他知道在还没有开口说话之时，销售就已经成功了一半。

掠夺式广告商的策略就类似于此，但掠夺式广告的规模更为庞大，其每天可以锁定数百万的潜在客户。当然，客户的无知是这场骗局的关键。营利性大学的许多目标学生都是外国移民，这些人相信，私立大学比公立大学更好。如果他们说的私立大学是指哈佛大学或者普林斯顿大学，那么这种观点是没错的。但是，所有的私立大学，包括像德瑞大学、菲尼克斯大学这样的营利性大学都比任何州立大学（比如伯克利大学、密歇根大学或弗吉尼亚大学）更好的说法，只有新来者才会相信。

一旦无知的特征被确证，招生人员接下来的主要工作就是锁定其中最易受欺骗的那些，然后利用他们的私人信息对他们展开针对性的“攻击”，就像江湖骗子那样。这就需要招生人员找到目标群体的最痛苦之处，也就是所谓的“痛点”，可以是低自尊，可以是在治安极差的贫民区单身抚养孩子的压力，也可以是药物成瘾。很多人在谷歌搜索引擎上搜索其困惑的问题或者在填写关于大学申请的调查问卷时，就不知不觉地暴露了自己的痛点。掌握了这一关键信息，招生人员就可以向这些人承诺，他们的大学所提供的高价教育能帮他们解决这些问题，消除痛苦。“我们只和得过且过、没有未来的人打交道，”翡特罗特学院在面向招生人员发放的培训材料中这样解释，“他们做出是开始上课、留在学校还是干脆辍学的决定，更多的是基于情绪而不是逻辑。对于过一天是一天的人来说，痛苦是他们最大的驱动力。”艾梯理工学院的一个招生团队甚至为此画了一幅宣传海报，其上一名牙医正在折磨着他痛苦的病人，旁边的标语则写着“找到他们的痛点”。

一个潜在客户第一次点击一所营利性大学的广告，其背后是一整个产业链的铺垫。例如，科林斯大学有一个30人的营销团队，每年有活动经费约1.2亿美元，他们会把大部分钱花在挖掘和追踪240万个潜在客户上，经过一年的努力工作，团队最终招到6万名新生，年收入高达6亿美元。这些大型营销团队通过各种渠道接触潜在的学生客户，包括电视广告、公路广告牌、广告邮件以及谷歌的搜索广告，甚至是前往学生所在的学校或住所直接拜访。团队中的分析师负责设计各种各样的促销活动，以获取反馈为第一目标。为了优化招生途径，也为了提高收入，他们需要知道他们投放的广告信息的接受者是谁，以及如果可能的话，这些信息对他们有什么影响。只有掌握这些数据，他们才能进一步优化招生途径。

当然，任何优化方案的关键都是确定一个目标。对于像菲尼克斯大学这样的文凭工厂，我认为可以肯定的是，它们的目标是尽量多地招收有资格申请政府助学贷款的学生。为了达成目标，团队中的数据科学家必须弄清楚如何才能更好地利用他们的各种营销渠道，使得他们花出去每一分钱都能产生最大的效益。

数据科学家首先使用贝叶斯方法优化营销渠道，这种方法从统计学的角度来看非常接近普通债券。贝叶斯分析的要点是根据变量对期望结果的贡献程度对变量进行排序。分析师会根据每一美元投入所产生的收益来评估搜索广告、电视广告、广告牌和其他广告形式的宣传效率。在评估结果中每种广告形式都有一个对应的体现其对收益的贡献程度的得分（以概率表示），这个概率将被用于调整下一次广告投放中该广告形式的权重。

这个评估过程很复杂，因为各种形式的推送广告会相互影响，而且其中的大部分影响都是无法测量的。例如，公共汽车广告能提高一个潜在客户拨打招生热线的概率吗？很难说。在线广告的效果追踪起

来更容易，而且商家可以用在线广告收集每个潜在客户的关键细节信息，比如他们的住址以及他们浏览过的所有网页。

这就是为什么营利性大学的大部分广告资金都流向了谷歌和脸书。广告商可以在每个平台上根据各种细节信息筛选目标人群。例如，贾德·阿帕图新电影的宣传人员可能会把目标观众定位在50个富人街区中的18~28岁的男性，锁定其中在脸书上点击过阿帕图热门电影《生活残骸》相关页面链接的人，在推特上提到过贾德或者其朋友提到过贾德的人，针对这些人投放电影广告。营利性大学的目标人群则与之相反，其更倾向于锁定贫困地区的人，尤其关注那些点击了发薪日贷款广告或者关注创伤后应激障碍问题的人。（退伍军人是营利性大学最主要的生源，某种程度上是因为他们更容易得到资助。）

营利性大学会设计一系列互相竞争的广告，考察哪种广告设计能带来最多的潜在客户。这是直邮营销从业人员几十年来一直使用的所谓A/B测试法的现代版本。直邮营销人员会发送大量的广告信息，评估受众对不同广告的不同反应，并据此调整广告设计。每当你又一次在邮箱里发现办理信用卡的广告邮件时，你就是在参加一次A/B测试。如果你看都不看就直接把邮件扔掉，你就给广告商提供了一个有价值的数据：这条广告对你没有价值。于是下一次，他们就会尝试一种稍微不同的方法再次发送邮件。这种做法似乎是徒劳的，因为那么多的广告邮件最终都被扔进了垃圾箱。但对于很多直邮营销人员来说，不管他们是通过网页广告形式还是邮件形式投放广告，能得到1%的回应率就已经是成功了。毕竟，他们掌握的数据是海量的。美国总人口的1%就已经是300多万人了。

这些营销活动一经迁移到互联网，商家对客户的了解就立即得到了加速。互联网为广告商提供了有史以来规模最大的消费者研究及客户开发实验室。网络广告在投放后的几秒之内就会得到反馈，比邮件快多了。在几小时（而不是几个月）之内，网络广告商就能够确定最

有效的投放方式，以极快的速度接近整个广告业的理想目标：让广告在对的时间出现在潜在客户眼前，以精准的信息触发顾客的消费决定，从而成功地捕获又一个付费客户。这种“投放—微调”式的广告营销活动永远不会停止。

随着时间推移和技术进步，基于大数据建立的模型开始能够自行筛选关于我们的数据，从中找到我们的习惯、意愿、忧虑和渴望。随着人工智能的机器学习能力快速成长，计算机只需要被赋予几条最基本的指令就可以最大限度地挖掘数据。算法能自行发现模式，然后根据时间变量将它们和某个结果联系起来。从某种意义上来说，计算机的确是在学习。

比起人脑学习，机器学习的效率并不高。一个孩子不小心碰到了炉子，感到疼痛，从这以后，她就会知道热金属和她快速收回手这一动作之间的关系。她还学会了一个词：烫。相反，机器学习程序需要大量的数据点才能创建其因果统计模型。但是现在，开天辟地第一次，海量数据变得很容易获得，同时拥有强大数据处理能力的计算机也被发明出来了。对于很多工作来说，机器学习比起受规则约束的传统程序更灵活、更精密。

比如，语言学家从20世纪60年代到21世纪初，一直在尝试着教计算机学会阅读。在此期间，他们的主要工作是对词义和语法规则进行编码。但是，每个初学外语的学生都能很快发现，语言中存在大量的例外，我们有俚语和讽刺，而且有些词的意思会随着时间和地理位置的变化而改变。复杂的语言是编程者的噩梦。语言编码在当时似乎毫无成功的希望。

但是，由于互联网的发明，世界各地的人们在生活、工作、购物和友谊等方面创造了海量新词。我们无意间为自然语句2机器建立了有史以来最大的训练语料库。当我们从在纸上书写语言转向在电子邮件和社交网络中敲打字母后，机器得到了学习自然语言的途径，其得以

将一种语言与其他语言进行比较，收集各类语境信息。机器学习的速度快得惊人。2011年，大多数科技公司还尚未看到苹果公司开发的可以识别自然语言的“个人助手”Siri的潜力。当时，Siri只精通某些领域的词汇，而且会犯极为可笑的错误。我认识的大多数人在当时都觉得它可以说是毫无用处。但现在我经常听到人们和他们的手机“交谈”，询问天气预报、体育比赛成绩、目的地的方位。2008~2015年，大概就在这段时间里，算法的语言能力从幼儿园水平提高到了中学生水平，其中一部分应用程序的语言能力进步得还要更快。

自然语言识别算法的进步为广告商创造了丰富的可能性。现在，程序能“知道”一个词的意思，或者至少“知道”可以把这个词与哪些行为和结果联系起来了。受助于此，广告商得以探索人的更深层次的行为模式。一个广告程序仍可以先从对人口和地理信息的筛选开始，但在几周或者几个月的时间里，这个广告程序就会开始学习目标人群的行为模式，对他们的下一步行动做出预测。广告程序会逐步地对目标人群有越来越深刻的理解。如果这是个掠夺式的广告程序，它就会评估目标人群的缺点和弱点，并找到最有效的途径榨取他们的价值。

除了最前沿的计算机科学，掠夺式广告商还常常与中间商合作，后者使用更原始的方法挖掘潜在客户。2010年，一条引人注目的广告便使用了一张奥巴马总统的照片，广告词为：“奥巴马呼吁妈妈们回到学校：完成你的学位——政府资助更偏爱有文凭的人。”这则广告暗示，总统签署了一项旨在让母亲重返校园的新法案。这显然是一个谎言。但如果人们受到总统照片的吸引点击了广告，这条广告就达到了目的。

此类误导性的广告标题，是一整个肮脏行业的劳动成果。根据独立组织ProPublica的调查，当一个消费者点击了这个广告之后，她会被问及几个问题，包括她的年龄和电话号码，然后一所营利性大学的招生人员就会立即联系她。给她打电话的人并不会给她提供更多关于奥

巴马总统新法案的信息——因为这项法案本就是无稽之谈。相反，他们会向她推销帮她借钱上大学的服务。

这种利用互联网筛选目标人群的程序被称为“潜在客户开发”程序。它的目标是列出一份潜在客户的清单，中间商会将其出售给相应的商家，在上述案例中就是卖给营利性大学。根据ProPublica的报告，营利性大学20%~30%的广告预算都用于潜在客户开发程序。大学每招到一个学生，就会支付给开发者150美元的中介费用。

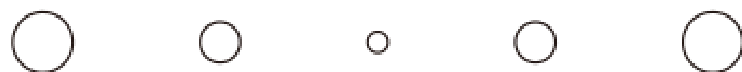
据公共政策研究员大卫·霍尔柏林透露，位于美国盐湖城的潜在客户开发公司中子互动（Neutron Interactive）会在Monster.com等招聘网站上发布虚假职位，还会发布一些承诺帮助人们获得食品券和医疗补助的广告。营销人员使用了同样的优化方法，即推出大量不同形式的广告，衡量这些广告对每个人的有效性。

这些广告的目的是诱导那些急于找工作的人提供他们的手机号码。在得到电话号码并与他们取得联系之后，大约有5%的人会对大学课程产生兴趣。尽管这个概率很低，但这些人是珍贵的潜在客户。营利性大学要为每个名字支付高达85美元的费用，而它们会千方百计地确保这笔投资换来丰厚的回报。据美国政府问责局的一份报告，在注册大学网站的5分钟之内，潜在的学生客户就会接到招生电话，注册后的一个月内，每个潜在的学生客户平均会接到180多个招生电话。

当然，营利性大学也有开发潜在客户的独特方式。它们最得力的工具之一是美国大学理事会网站，许多学生在这个网站上注册报名SAT考试，研究人生的下一步要怎么走。布鲁克林的一所公立学校，女子数学科学城市学院（Urban Assembly Institute of Math and Science for Young Women）的入学顾问马拉·塔克表示，贫困学生在搜索引擎中看到的关于大学申请的搜索结果将引导他们申请营利性大学。一旦有学生在填写某个网络调查问卷时表示她需要助学金，营利性大学的网站就会出现在搜索结果中匹配学校列表的前列。

营利性大学也提供一些免费的服务，用以换取与学生面对面交谈的时间。卡西·麦格西斯也是一名入学顾问，他告诉我，这些大学会提供免费的培训指导学生们写简历。这些培训确实对学生有益。但是，其中那些提供了联系方式的贫穷学生在培训结束后就会成为大学的追踪目标。营利性大学不会去打扰富人学生，因为这些学生和他们的父母有很强的鉴别能力。

以各种形式招收学生是营利性大学的核心业务，在大多数情况下，它们在这方面的资金投入要比在教育上的投入多得多。参议院对30所营利性大学所做的一项调查发现，从人数比例来讲，每48个学生就对应一个招生人员。菲尼克斯大学的母公司太阳神教育集团，其2010年在市场营销上的投入超过10亿美元，这些钱几乎都用在了招生上。花在每个学生身上的平均营销费用为2225美元，而平均教育费用则仅有892美元。相比较而言，俄勒冈州的波特兰社区大学对于每个学生的教育投入平均为5953美元，而平摊在每个学生身上的营销费用为185美元，仅占其预算的1.2%。



数学，以复杂模型的形式助长了整个掠夺式广告产业的发展，引导潜在的学生客户申请这些营利性大学。但当招生人员开始不断地打电话给这些学生的时候，我们便遗忘了幕后的数学模型。招生人员承诺合理的学费、光明的职业前景以及突破阶层的机会，这些说辞与那些灵丹妙药、脱发治疗和能减少腰围的振动带等广告没有什么不同，不是什么新鲜招数。

然而，数学杀伤性武器的典型特征之一就是它会危害很多人的生活。而这些受助于数学模型的掠夺式广告的危害直到学生们为了缴纳学费申请大额贷款时才会显现。

教育产业中，一个极为重要的参数是1965年颁布的高等教育法案中的90/10法规。该法规规定，学生申请的联邦助学贷款不能超过学费的90%。政府认为，只要学生为上大学有所支出，他们就会更认真地学习。但是营利性大学很快找到了钻空子的方法。他们给予学生的“优惠”政策是，如果学生可以依靠家庭储蓄或借助银行贷款凑到几千美元，那么营利性大学就可以将他们的贷款额度提高到政府贷款额度的9倍。这一政策让每个学生都能为学校带来极大的回报。

对许多学生来说，这项贷款听起来就像是免费送钱给他们，而学校则对这种误解乐见其成。但这并不是免费的钱，而是债务，很多学生很快就发现自己陷入巨额债务当中。破产的科林斯大学的学生的未偿还债务总计达35亿美元。这些钱几乎都来自纳税人，而他们永远不会得到偿还。

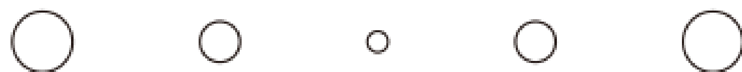
一些学生毫不犹豫地选择了营利性大学，并且在那里学到了对他们而言足够的知识和专业技能。但是，他们的职场青睐度真的会比学费低得多的社区大学的毕业生更好吗？2014年，美国研究协会的调查人员出于研究目的制作了近9000份虚拟简历。这些虚构的求职者中，一些人持有营利性大学的学位，还有些则是从社区大学那里获得了类似的文凭，而另外1/3的人则根本没有接受过大学教育。研究人员把这些简历投到7个大城市的招聘市场，然后统计了不同类型的简历的回应率。他们发现，营利性大学的文凭在职场中的价值要低于社区大学的文凭，和高中文凭相差无几。然而，这些大学的平均学费要比一流公立大学还高出20%。

这一数学杀伤性武器造成的恶性循环并不复杂。美国最贫困的40%的人口深处绝望。许多工业岗位消失了，要么被新技术取代，要么被转往海外。工会已经失去了影响力。20%的人口控制着全美89%的财富，而位于社会底层的40%的人口几乎一无所有。他们的资产是负的：在这个庞大的、在温饱线上苦苦挣扎的下层社会，普通家庭的

净负债为14800美元，其中大部分欠款来自利率过高的信用卡账户。这些人真正需要的是钱。而他们一直被灌输这样的理念，即能否赚钱的关键取决于接受教育的水平。

营利性大学使用的数学杀伤性武器精准度极高，直指最为弱势的群体。它们给弱势群体提供了一个接受教育的机会，并向他们承诺了一个美好的未来、一个诱人的突破固有阶层的机会，而结果是这些贫困家庭的学生因此背负了更多的债务。它们利用贫困家庭的迫切需求，利用他们的无知和殷切希望榨取利润，并且是大规模地榨取。这些贫穷的学生因此变得无望、绝望，怀疑教育的价值，这一切又进一步加剧了美国社会的贫富差距。

我想指出的是，这些文凭工厂是从“贫”和“富”两个方向上塑造并加剧了社会不公的。成功的营利性大学的校长每年能赚几百万美元。例如，美国菲尼克斯大学的母公司阿波罗教育集团的首席执行官格雷戈里·W. 卡培里在2011年的薪酬就高达2510万美元。在公立大学里，只有足球教练和篮球教练才有可能拿到这么高的薪酬。



不幸的是，营利性大学远非唯一利用掠夺式广告赚取巨额利润的机构，类似的机构和行业还有很多。只需想一下人们的痛点会出现在哪里，你就会在相应的领域发现广告商在大肆使用掠夺式模型。最大的商机之一就是贷款。每个人都需要钱，但有些人的需求比其他人更迫切。这些人不难找到，他们很有可能居住在贫困地区。站在掠夺式广告商的角度，穷人在搜索引擎上的查询行为和点击过的优惠券，实际上就是在大声呼叫广告商来注意自己。

和营利性大学一样，发薪日贷款行业也在操纵着数学杀伤性武器。虽然其中一些操纵者是合法的正规公司，但这个行业在本质上就

是掠夺性的：通过设置高得离谱的利率——平均利率高达574%——赚取暴利。发薪日贷款的还款额平均为借款额的8倍，从这一指标衡量，其更像是长期贷款。发薪日贷款行业得到大批数据代理商和潜在顾客开发商的大力支持，这些人中的很多人都可以被称为“诈骗艺术家”。他们设计并发布的广告会突然出现在你的电脑屏幕和手机屏幕上，以提供方便、快捷的获取现金的机会吸引你点击。潜在客户在点击后填写的申请表中往往会写下其银行账户信息，这让他们主动成为欺骗和信息滥用的受害者。

2015年，美国联邦贸易委员会指控两家数据代理商非法销售50多万名消费者的贷款申请信息。根据起诉书，这些公司，包括位于佛罗里达州坦帕市的红杉资本（Sequoia One of Tampa）和克利尔沃特市的X世代营销集团（Gen X Marketing Group），窃取客户贷款申请中的电话号码、雇主信息、社会保险号和银行账户信息，然后以每条信息约50美分的价格出售。据监管机构统计，购买这些客户信息的公司至少给这些客户造成了总计710万美元的损失。许多受害者还因账户被清空或无效支票而被银行收取了各项手续费。

如果你想要知道客户的具体损失数额，我可以告诉你，数额低得可怜。710万美元的损失分摊到50万个账户上，每个账户平均仅损失了14美元，但他们被盗取的钱很可能是这些贫困家庭的银行账户里的最后50美元或100美元。

现在，监管机构正在推动个人信息数据市场的立法，个人信息数据是所有数学杀伤性武器的关键参数。到目前为止，政府已发布了一些联邦法律旨在对个人医疗和信用信息数据进行保护，如《公平信用报告法》和《医疗保险可携性和责任法案》（HIPAA）。也许，我们可以期待联邦政府在深挖潜在客户开发商的种种恶劣行径之后，能推动更多的个人信息数据方面的立法。

然而，我们将在下一章中看到，一些最高效，同时也是最恶劣的数学杀伤性武器能绕过法律和监管。它们研究有关我们的一切——从我们生活的社区到脸书朋友圈——以预测我们的行为，甚至压缩我们的生活空间。

第五章
效率权衡与逻辑漏洞

大数据时代的正义



后工业时代，宾夕法尼亚州的小城雷丁经历了一段艰难的时期。雷丁坐落于费城以西50英里的绿色山丘上，是靠铁路、钢铁、煤炭和纺织业发展起来的。但最近几十年，随着这些行业的急剧萎缩，雷丁城也走向衰落。到2011年，该地区成为全美贫困率最高的地区，贫困率高达41.3%。（不过第二年，其倒数第一的名次就被底特律抢走了）。在2008年的经济危机之后，经济衰退重创了雷丁的经济，税收收入的下降，导致尽管当地犯罪事件不断，但还是有45名警察被裁掉了。

雷丁的警察局局长威廉·海姆不得不想办法以更少的警力维持同等或更好的治安环境。因此，2013年，他引进了位于加州圣克鲁斯的大数据分析初创公司PredPol所开发的犯罪预测软件。该软件根据地区的历史犯罪数据，按小时计算最可能发生犯罪的地点。该软件预测出的

犯罪地点可被看作一块相当于两个足球场那么大的正方形区域。如果警方在相应的时间在这些预测区域安排更多警力巡逻，那么他们就很有可能阻止犯罪的发生。果然，一年后，海姆警长宣布，雷丁的盗窃率下降了23%。

类似PredPol所开发的这种预测软件在全美各地预算紧张的警察局那里得到了广泛的应用。亚特兰大和洛杉矶的警察局同样根据预测实时变换警力部署，两地都报告了犯罪率在下降。纽约市的警察局也使用了类似的软件，叫作CompStat。费城警局则使用了一款名为“HunchLab”的由本地公司开发的软件，该软件的功能还包括对犯罪高发地区的分析，自动取款机或便利店等极易发生犯罪的特殊地点都被计算在内。和其他大数据行业一样，犯罪预测软件的开发人员也忙于搜集所有能提高其模型的预测准确性的信息。

仔细考虑一下你就会发现，针对犯罪行为开发的预测软件与我们之前讨论过的用于变换防御阵型的棒球模型类似。棒球模型通过研究每个球员的击球历史数据，安排外野守在可能性最高的落球点接球。犯罪预测软件也进行了类似的分析，并据此安排警察前往最有可能发生犯罪的地方巡逻。两种模型都致力于优化资源配置。但大部分犯罪预测模型更为复杂，因为它们还需要预测下一次犯罪高发期的起始点。例如，PredPol所开发的犯罪预测软件其原理就借鉴自地震预测软件：研究一个地区的犯罪发生情况，将其整合到已有的历史模式中，以此预测下一次该地区的犯罪事件可能发生的时间和地点。（PredPol发现了一个简单的关联：如果窃贼袭击了你隔壁邻居的房子，那你也要做好准备了。）

像PredPol所开发的这种犯罪预测模型的确有其自身的优点。不同于史蒂芬·斯皮尔伯格的反乌托邦电影《少数派报告》（以及现实世界中一些不好的预兆，这一点我们将在后文简单阐述），警察不会追踪尚未犯罪的人。杰弗里·布兰丁汉姆是加州大学洛杉矶分校的人类学教

授，也是PredPol公司的创始人，他对我强调，这种模型没有将种族和民族作为指标纳入其中。跟我们第一章讨论的再犯风险评估模型之类的软件被用于判决指导不同，PredPol所开发的软件模型并不关注个人，它关注的是地理位置。关键输入信息是每一次犯罪事件的类型、地点以及发生的时间。这看起来很合理。如果根据预测结果安排警察多在犯罪高发区巡逻，阻止盗窃抢劫事件的发生，我们完全可以相信社区会因此受益。

但是大多数犯罪并不像入室盗窃和偷盗汽车那样严重，这也是模型出现问题的地方。当警方采用PredPol的预测模型时，他们需要做一个选择。他们可以只关注所谓的第1类犯罪，即暴力犯罪，这些犯罪事件通常都是必须被警方记录在案的，包括杀人、纵火和人身侵犯。或者，他们也可以扩大关注范围，将第2类犯罪也囊括在内，如流浪、侵犯型乞讨、贩卖和消费少量毒品等。如果不是警察就在现场，这种轻微的妨害罪通常不会被记录在案。

这些轻微的妨害罪在许多贫困社区都很常见。在有些地方，警察称之为“反社会行为”，简称ASB。不幸的是，在模型中囊括这些妨害罪的数据会导致预测结果发生巨大偏差。一旦将这些妨害罪的数据计入犯罪预测模型，就会有更多的警察被派进贫困社区，而警察在这些社区很有可能会逮捕更多的人。毕竟，即使警察的主要目标是阻止盗窃、谋杀和强奸，他们也一定会有相对空闲的时段。巡逻的性质就是如此。而如果一名正在巡逻的警察看到了几个看起来不超过16岁的孩子在大口大口地喝着装在棕色纸袋里的瓶装酒，他就会去阻止这些孩子。当这些轻微犯罪被更多地记录在案时，警方的犯罪预测模型中就会出现越来越多的此类数据点，据此，警察将被再次派往这些轻微犯罪高发的贫困社区。

这就产生了一个恶性循环。警方自身不断制造新数据，证明了对贫困社区加大警力的合理性。而美国监狱里则挤满了成千上万个妨害

罪罪犯。他们大多数来自贫困社区，很多是黑人或拉美族裔。也就是说，即使这类模型不关心种族因素，其结果仍然体现了种族偏差。在存在种族隔离的城市中，地理位置是种族的一个非常有效的替代变量。

你可能会问，如果预测模型的目的是防止重大犯罪，那为什么要追踪那些轻微犯罪呢？犯罪学家乔治·凯林与公共政策专家詹姆斯·威尔逊于1982年在《大西洋月刊》上发表的关于"破窗理论与社会治安"的文章对此问题进行了探讨。自这篇文章发表以来，人们便坚信，反社会行为与犯罪之间存在必然联系。这篇文章认为，低级犯罪和轻罪造就了一个混乱的社区氛围，吓走了守法的公民，留下了黑暗和空荡的街道，而这些地方正是滋生严重犯罪的温床。应对办法是阻止无序的蔓延，包括修复破碎的窗户，清理涂满涂鸦的地铁车厢，采取措施阻止轻微犯罪等。

破窗理论引发了20世纪90年代各地的"零容忍"运动热潮，以纽约市为最。纽约市的警察会因为地铁逃票而逮捕孩子，甚至连在场的逃票者的朋友也全都不放过，警车会在城市里轰隆隆地行驶几个小时，直到把他们全部逮捕归案。有些人认为，警方大刀斧阔的行动能极大地减少暴力犯罪，另一些人则不同意这种看法。还有一些人甚至把犯罪率的下降与20世纪70年代推行的堕胎合法化联系在一起。许多其他的与之类似的理论也纷纷出炉，包括将可卡因成瘾人数的下降和20世纪90年代的经济繁荣联系在一起。不管怎样，零容忍运动获得了广泛的支持，刑事司法系统将数百万个年轻的少数族裔男子送进监狱，而其中的许多人只犯了轻罪。

但是，零容忍实际上与凯林和威尔逊的破窗理论没有什么关系。他们的案例研究的是新泽西州纽瓦克市的一次成功的治安行动。按照他们的理论，在治安较差的贫困社区巡逻的警察应该以高度宽容的态度管理社区。他们的工作是去适应社区自身的秩序标准，并帮助社区

维持这个秩序标准。一个城市的不同社区，其秩序标准也各不相同。有的社区要求酒鬼们必须把酒瓶装在袋子里，避开主要街道，但在小巷则没有这个要求。瘾君子可以坐在楼前台阶上，但不可以就地躺倒。研究者强调的是维持标准，只要求警察帮助社区维持其自身秩序，而非强行施加另行设立的管制。

你可能会觉得话题已经偏离了PredPol的软件、数学以及数学杀伤性武器。但是，破窗理论和零容忍运动也都分别代表了一类模型。就像我的家庭饮食模型或者《美国新闻》的大学排名模型一样，每一个打击犯罪的模型都需要一些输入数据，并会给出一系列的反馈结果，每一个模型也都有其特定的目标。用这种思维研究我们所讨论的治安工作很重要，因为现在，正是这些数学模型主导着现实层面的执法行动，而其中的一些模型就是数学杀伤性武器。

回到之前的话题，现在我们可以理解警察局为什么要把轻微犯罪的数据纳入PredPol的软件中了。生活在以零容忍运动为正统的社会环境中，很多人几乎从来不曾怀疑轻罪与重罪之间的必然关系，认为这就像烟会导致火一样。英国肯特郡的警察局在2013年引入了PredPol的软件，那里的警察也将轻微犯罪的数据纳入了预测模型。模型看起来确实奏效了。他们发现，根据PredPol软件的指引，在指定区域巡逻逮捕罪犯的效率是随机巡逻的10倍，预测精确度是警方情报分析的2倍。那么该模型最擅长预测的是什么类型的犯罪呢？轻微犯罪。实际上，在世界各地应用该类模型都将导向这个结果。醉汉总会在同一面墙上尿尿，瘾君子总会睡在同一个公园的长椅上，而偷车贼或抢劫犯则总会四处走动，试图预测警察的动向。

即便警察局局长强调要以打击暴力犯罪为主，避免让大量的轻微犯罪数据被纳入犯罪预测模型也需要极大的克制力。人们很容易相信，数据越多越好。只关注暴力犯罪的模型，其统计图像上可能只有

零星几个点，而纳入轻微犯罪数据的模型则会“完美”地呈现出一个治安混乱的城市面貌。

可悲的是，在大多数地区，这样的犯罪地图将与该地区的贫富分布图大幅重叠，而由此得出的贫困地区犯罪率高的结论只能证实并强化社会中上层阶级的普遍观点：穷人要为自己的境遇负责，大部分坏事都是穷人干的。

但是，如果警察去追踪不同类型的犯罪呢？这听起来可能有些违反直觉，因为包括警察在内的大多数人都把罪行看作一个金字塔。塔顶是谋杀，往下是稍常见一些的强奸和人身侵犯，然后是几乎天天发生的商店行窃、小额欺诈，甚至违规停车。把谋杀罪放在犯罪金字塔的顶端是合理的。尽可能地降低暴力犯罪是警方最核心的任务，这一点相信绝大多数人都会同意。

那么，在PredPol软件所呈现的犯罪地图上，那些富人实施的犯罪行为又是怎样的情形呢？在21世纪的头10年，金融大亨夜夜狂欢。他们满嘴谎言，用数十亿美元对赌客户的投资，欺诈，贿赂评级机构。他们犯下了滔天罪行，在21世纪的最初5年摧毁了全球经济。数以百万计的人因为他们失去了自己的家园、工作和医疗保障。

我们完全有理由相信，如今的金融领域有更多这类犯罪行为。如果我们学到了什么，那就是：金融界的驱动目标是赚取巨额利润——越多越好，以及，任何类似促进自我调节或自我纠正的做法在该领域都是毫无意义的。由于金融界拥有巨额财富和强大的游说集团，金融行业在很大程度上并不处在监管之下。

试想一下，警方在金融领域实施零容忍策略将会如何。他们会因为哪怕是最轻微的违规行为而逮捕从业者，无论是欺骗投资者加入401k计划，提供误导性的投资建议，还是小额欺诈。也许特警队会突

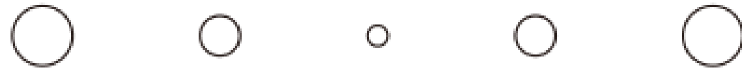
然出现在康涅狄格州的格林尼治，在芝加哥商业交易所附近的小酒馆里开展卧底行动。

当然，这是不可能的。警察不具备开展这种行动的专业技能。他们所有的培训、装备，从上岗训练到防弹背心，都只适用于小街小巷。打击白领犯罪需要的是完全不同的工具和专业人才。美国联邦调查局和证券交易委员会的调查组，这些负责打击此类罪行的资金不足的小团队从几十年前就知道，银行家实际上是无懈可击的。他们在国家政要身上投入了大量资金，并且往往会从后者那里得到相应的回报，同时，他们的存在也被认为对于国家整体经济体系正常运转而言至关重要。他们因此被全方位地保护起来了。如果他们经营的银行日渐衰退，国家整体经济水平也会随之下降。（穷人就没有这样有力的申辩理由。）因此，除了一些像庞氏骗局操作者伯纳德·马多夫这样的极端罪犯，金融家是不会被逮捕的。这个群体在2008年的市场崩盘中几乎毫发无损，我们现在又能拿他们怎么样?

我真正想说的是，警察可以主动选择把关注点放在哪儿。今天，他们几乎只关注穷人。这是他们的一贯作风，也是他们的使命所在——至少他们自己是这么认为的。而现在，数据科学家正在将这种社会秩序的现状与PredPol之类的预测模型进行对接，这类模型将对我们的生活产生巨大的影响。

其结果是，尽管PredPol所开发的是一个非常有益，甚至可以说是高尚的软件工具，但它同时也是一个会自行发展的数学杀伤性武器。从这个意义上来说，即使是出于最高尚的初衷，PredPol的预测软件仍然授权了警察局锁定穷人、拦截穷人、逮捕穷人，并把其中的一部分穷人罪犯关进监狱。警察局长，如果不是全部，至少也是大部分，则认为他们正在采取唯一合理的途径来打击犯罪。他们指着地图上的贫困社区说，这就是犯罪高发地。先进的大数据技术为他们一贯以来的观念赋予了精确性和“科学性”。

结果是，我们一边将贫困视为犯罪，一边相信我们的工具既科学又公平。



2011年春天的一个周末，我参加了纽约市一个数据分析方面的“黑客马拉松”活动。这类活动的目标是将黑客、电脑迷、数学家和软件极客们聚集在一起，动员这个高智商群体阐释对我们生活产生巨大影响的数字系统。我和纽约公民自由联盟的小队结成了搭档，我们的工作就是要“攻破”纽约警察局实行的旨在预防犯罪的重大行动之一，即所谓的“拦截、审问、搜身”行动。大多数人只知道这项行动被简称为“拦截—搜身”，但值得引起注意的是，在数据驱动的时代里，此类行动的开展频率和规模急剧增加。

警方认为，拦截—搜身行动是一种犯罪过滤器。道理很简单，警察会拦截那些看起来可疑的人——可能是因为他们走路或穿衣服的方式，也可能是因为他们的文身，然后找他们问话，或问也不问就把他们按在墙上或汽车的引擎盖上。警察会向其索要身份证，然后进行搜身。人们认为，如果警察拦截了足够多的人，这毫无疑问能阻止大量的轻微犯罪，也许还包括一些重大犯罪。这项政策由当时的纽约市市长迈克尔·布隆伯格推行实施，得到了公众的大力支持。在过去的10年里，警方实行拦截行动的次数增加了6倍，达到了近70万次。绝大多数被拦截的人都是无辜的，而这一遭遇令他们非常不愉快，甚至极为恼火。然而，公众将这一政策与纽约市犯罪率的急剧下降联系了起来。许多人认为纽约更安全了。统计数据也显示出同样的结果，谋杀案件由1990年的2245件下降到515件（到2014年又下降到400件以下）。

每个人都知道，在警察拦截的人中，有一大部分是年轻的黑人。但是警察具体拦截了多少人？这些被拦截的人被证实涉嫌违法并被逮捕的概率，或者拦截行动成功阻止犯罪的概率是多少？尽管这方面的

信息根据法律规定应该是公开的，但其中很多数据都存储在非机关人员无法访问的数据库中。我们在这次“黑客马拉松”活动里的任务就是攻破这个数据库，释放其中的数据，这样我们就可以分析“拦截—搜身”政策的本质和效力。

结果是，我们发现，大约85%被拦截的人是非洲裔或拉美裔的年轻人，对此发现我们并没有特别惊讶。在某些社区，这两类少数族裔群体中有很多人都曾被多次拦截。而被拦截的人中只有0.1%，即1000个人里只有1个与暴力犯罪有关系。同时，利用此犯罪过滤器，警方捕获了更多涉嫌轻微犯罪的人，包括吸毒者和饮酒的未成年人。此外，如你所料，一些被拦截的无辜民众由于对此遭遇感到过于愤怒而进行了激烈反抗，这些人最终反倒因此而被指控拒捕。

纽约公民自由联盟对布隆伯格政府提起诉讼，指控其“拦截—搜身”政策有种族歧视倾向，是不公平的治安管理行动的典型实例，该政策把更多的少数族裔推进刑事司法系统，送进监狱。联盟指出，黑人男性入狱的可能性是白人男性的6倍，被警察击毙的可能性是白人的21倍，至少根据现有的数据（显著低报的数字）是如此。

“拦截—搜身”并不完全是一种数学杀伤性武器，因为该政策依赖于人的判断，并没有被程序化，但此政策建立在一个简单却极具破坏性的算法之上。如果警察在某些社区拦截了1000人，他们在其中平均会发现一个重大犯罪嫌疑人和许多的轻微犯罪嫌疑人。这与掠夺式广告商或直邮营销的目标回应率没有太大区别。如果基数足够大，即使命中率很小，你仍然能实现你的目标。这也解释了为什么在布隆伯格政府的大力推行下，“拦截—搜身”行动的规模增长得如此之快的原因。如果被拦截的人增加6倍，被逮捕的人也能随之增加6倍，那么成千上万无辜的人所遭受的不便和骚扰就会被认为是合理的牺牲——难道他们不想阻止犯罪吗?

但是，“拦截—搜身”政策在很多方面也与数学杀伤性武器有相似之处。例如，它产生了恶性循环，导致成千上万的黑人和拉美裔被逮捕，而其中的许多人只是犯了点儿轻罪或者有违规行为——这些轻微罪行在每个周六的晚上都会在大学兄弟会发生，而大学生却没有因此受到惩罚。犯下同样的轻微罪行，绝大多数大学生毫发无损，而“拦截—搜身”政策的主要受害者——少数族裔却被警方记录在案，他们中的一些人甚至因此被发配到里克斯岛监狱。更糟糕的是，每一次逮捕都创造了新的数据，这又进一步证明了这一政策的正当性。

随着“拦截—搜身”行动的发展，把该政策当作预防犯罪的有效手段的合理根据逐渐丧失了其实际意义，因为警方不仅会追踪那些以前可能犯过罪的人，而且还会追捕那些未来可能会犯罪的人。毫无疑问，警察有时能够完成这个目标。警察可能因为逮捕了一名身上藏有未注册枪支的年轻男子，从而使这个社区免于一场潜在的谋杀或持枪抢劫，甚至连续抢劫。当然，也可能什么也不会发生。但不管怎样，“拦截—搜身”行动看起来都是有理有据的，许多人也认同该政策具有说服力。

但是，这一政策合乎宪法吗？2013年8月，联邦法官希拉·谢德玲裁定，该政策违反了宪法，她指出，警察经常“拦截黑人和拉美裔，如果他们是白人，就不会被拦截了”。她在判决书中写道：“‘拦截—搜身’政策违反了美国宪法第四修正案，该修正案保护民众免受政府不合理的搜查和扣押，而且‘拦截—搜身’政策也未能实现宪法第十四修正案所承诺的平等保护。”她呼吁对该政策进行广泛的改革，包括要求对巡逻警察的行动予以录像，这将有助于确定某次行动是否有合理根据的，并在一定程度上消除“拦截—搜身”模型的不透明性。不过，这项裁决对改善不公平的治安管理体制并没有产生根本性的影响。

在研究数学杀伤性武器的时候，我们常常需要在公平和效率之间进行权衡。我们的法律传统更倾向于公平。例如，宪法就假定一个人

是清白的。站在建模者的立场，无罪推定是一个约束条件，其带来的副作用让一些确实有罪的人被判无罪释放，特别是那些能够请得起优秀律师的人。即使是那些被判有罪的人也有权对判决提出上诉，而这又会消耗大量的时间和资源。因此，我们的法律体系在很大程度上牺牲了效率来保证公平。宪法的隐含判断是，相比监禁或处决一个无辜的人，因缺乏证据释放一个很可能犯了罪的人对我们的社会造成的危害更小。

相反，数学杀伤性武器更倾向于效率。本质上，数学杀伤性武器建基于可测量和可计算的数据。但公平是模糊的，很难量化，它是一个抽象概念。我们的计算机程序尽管在语言学习和逻辑学习方面有所进步，但仍然不能很好地理解抽象概念。它们所理解的“美”只是一个与大峡谷、海洋日落和时尚杂志的美容美发相关联的词，它们试图通过计算脸书上的点赞数和关系网来衡量“友谊”。而到目前为止，计算机还完全不理解公平这个概念。程序员不知道该如何为公平编码，他们的老板也很少会要求他们做这件事。

所以，公平的概念没能被编入数学杀伤性武器，这导致了大规模的、产业化的不公平。如果你把数学杀伤性武器想象成工厂，那么不公平就是那些从烟囱里冒出来的黑色排放物。这些黑色排放物是有毒的。

问题是，我们的社会是否愿意为了公平而牺牲一点儿效率呢？如果我们丢弃一些数据，给模型完成任务增加点儿难度会如何呢？例如，输入大量的反社会行为的数据可能有助于PredPol的软件描绘出一幅更精确的犯罪高发区地图，但代价是造成恶性循环，所以有理由认为，我们应该抛弃这部分数据。

这是一件很棘手的事情，在许多方面与针对美国国家安全局的电话窃听行为的抗争行动相似。此类窥探行为的支持者认为这对我们的安全很重要。为了完成使命，那些运行和维护这个庞大国家的安全保

障体系的人会不断地努力收集更多的信息。他们会继续侵犯人们的隐私，直到他们意识到，必须找到一种符合宪法的工作方法，否则公众的抗议将使其工作难以推进下去。抗议很难，但很有必要。

另一个问题是平等。如果我们现在说，为了公平，每个人包括穷人和富人，白人和黑人，都得忍受“拦截—搜身”行动的骚扰和屈辱，那么公众还会愿意这项牺牲了“合理根据”的行动继续开展下去吗？如果警方以公平的名义派一群警察去芝加哥繁华的黄金海岸巡逻会怎么样呢？芝加哥警察局有他们自己的一套“拦截—搜身”的方法。也许在黄金海岸，警察也会仅仅因为市民不遵守交通规则从公园慢跑穿过北方大道或者在湖滨大道上遛狗就逮捕他们。这次巡逻安排很可能会导致当地更多的醉驾司机被逮捕，也许还会揭露一些保险诈骗、家暴或敲诈勒索等罪行。偶尔，可能只是为了让那里的人“体验生活”，警察们会突然逮捕那些正在豪华游轮上享受的富人，把他们关押在巡逻车的后座，用手铐反铐起来，同时嘴里还骂着脏话。

而随着时间的推移，黄金海岸巡逻行动会产生新的数据，当地犯罪率的上升将招来更多的警察。而这无疑会引起越来越多的愤怒和对抗。我想象的是这样一个画面，一个并排停车的司机，回过头和拦下他的警察说话，拒绝从他的豪华跑车上下来，于是发现自己被指控拒捕。就这样，黄金海岸就又发生了一起犯罪事件。

这听起来可能不怎么严肃，但我想说的是，正义的关键是平等。这意味着，刑事司法行动也应该被包括在内。那些赞成“拦截—搜身”政策的人应该亲自体验一下这项行动。正义不能仅仅由社会的一部分人强加给另一部分。

被拦截、被指控、被逮捕归案并被记录到司法系统中，并不意味着“拦截—搜身”政策或PredPol的预测模型所造成的恶性循环就此结束。一旦被捕，许多人将会遭遇另一种数学杀伤性武器的袭击，即我在第一章中讨论的用于指导量刑的再犯模型。不公平的治安管理体系

所产生的不公正的数据现在都集中到了这一模型中。接下来，检察官会根据理论上更具科学性的模型分析，给每个罪犯一个风险等级分数。而重视这个分数的法官自然会给予更可能再犯罪的人以更长的刑期。

为什么来自贫困社区的少数族裔囚犯更有可能犯罪？根据再犯模型的数据，这是因为他们失业的可能性更高，没有高中文凭，有过前科，并且他们的朋友也有过前科。

但是，对于同样的数据还有另一种解读的方式，即这些囚犯住在贫困社区，社区内的学校很糟糕，机会也很少。而且他们受到警方更严格的监管。因此，比起刑满释放后回到富人区的金融诈骗犯，刑满释放后回到穷困社区的罪犯再次触犯法律的可能性也更高，这是毫无疑问的。在这个体系下，穷人和少数族裔因为他们的身份和出生地而遭受了更多的惩罚。

另外，再犯模型作为所谓的科学系统也存在逻辑漏洞。该模型假定把“高风险”囚犯关押更长的时间，社会就会更加安全。当然，在监狱服刑期间，犯人的确不会在社会犯下罪行。但是，一旦他们出狱，被关押的时间还会继续影响他们的行为吗？长期生活在一个被重罪犯包围的恶劣环境中，是否更有可能增加他们犯罪的概率，而不是减少呢？这一发现一旦被证实，将会直接削弱再犯判决指导原则的合理性。但是，拥有海量数据的监狱体系并没有进行这项非常重要的研究。他们只是在使用数据来反复地证明系统的合理性，而不会去质疑或改进系统。

我们可以将这种态度与亚马逊在类似情况中的态度进行比较。零售巨头亚马逊，就像刑事司法系统一样，也专注于开发一种“再犯”模型。但亚马逊的目标与后者恰恰相反——它希望人们反复来此网站购物。也就是说，亚马逊模型会锁定“再犯者”，并鼓励“再犯”。

现在，如果亚马逊像刑事司法系统那样运作的话，它首先会把消费者作为潜在的罪犯进行评分。也许“再犯”可能性高的用户中的大部分也生活在特定的地区，或者拥有大学学位。在这种情况下，亚马逊会更频繁地针对这些人投放广告，比如经常给他们提供购物折扣。如果该营销策略起效，那么这些有较高再犯分数的人就会重新购物。从表面上看，营销结果似乎证实了亚马逊评分系统的合理性。

但与刑事司法中的数学杀伤性武器不同，亚马逊并不满足于这种草率的相关性。亚马逊有一个数据实验室。如果公司希望查明什么是驱使消费者再次购物的因素，数据实验室就会着手对此问题进行研究。数据科学家不仅关注用户的地理位置和教育水平，还会考察其对于整个亚马逊销售平台的用户体验。他们可能会先研究那些在亚马逊购物过一两次而后来再也不来光顾的顾客的行为模式。他们在付款时遇到麻烦了吗？他们的包裹准时到达了吗？他们中有更多人给了差评吗？他们会不断地问问题，因为公司的未来取决于一个不断学习的系统，一个能找出客户需求的系统。

如果我有机会成为司法系统的数据科学家，我会尽我所能深入了解囚犯的内心世界，以及监禁经历对囚犯的行为有何影响。我首先会去研究单独监禁对囚犯的影响。每天都有成千上万名囚犯会被关押在一个监狱中的监狱长达23个小时，关押他们的禁闭室大多数都没有一个马棚大。心理学家已有研究发现，长时间的独处会让人产生绝望的感觉。那么，这一发现也适用于罪犯吗？这是我想要研究的一个问题，但是这方面的相关数据可能还没人采集过。

另外，监狱内部强奸事件的影响呢？在《不平等的审判》一书中，作者亚当·邦福拉多写道，某类囚犯在监狱里总会遭遇强奸。年轻人和身材矮小的人尤其容易受到侵犯，智力受损人士也一样。其中一些人在监狱中作为其他囚犯的性奴隶生活了很多年。这是另一个重要的可以分析的问题，对此，任何拥有相关数据和专业知识的人都能够

给出明确的解答，但监狱系统从来没有兴趣去记录此类虐待遭遇对囚犯的长期影响。

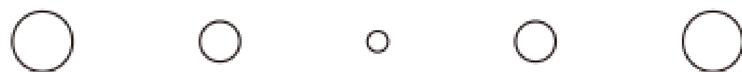
一个严谨的科学家也会从监狱经历中寻找积极的影响因素。更多的阳光、更多的运动、更好的食物、识字训练，这些因素会产生怎样的影响？也许这些因素会改善罪犯出狱后的行为。更有可能的是，它们会产生不同的影响。一个严谨的针对司法系统的研究计划将深入探究各个不同因素的影响，研究这些因素对囚犯的共同作用，以及这些因素对哪些人最有利。在相关数据有记录并被积极利用起来的前提下，我们就能更好地实现数据模型的目标，即为了囚犯和整个社会的利益而优化监狱，这与亚马逊之类的公司对网站或供应链进行优化的目的是一样的。

但是，监狱有充分的动机避免这些由数据驱动的研究。首先，监狱所面对的公关风险太大了，没有哪个城市愿意被《纽约时报》点名批评。其次，过度拥挤的监狱系统有巨额利润可图。私营监狱的市值高达50亿美元，而目前它们只容纳了囚犯总人口的10%。和航空公司一样，私营监狱在“客满”时才能获得更多的盈利。而太多的问题和研究可能会威胁这项收入来源。

因此，刑事司法系统不去分析监狱、优化监狱，而是把监狱隐藏起来，把它变成一个黑盒子。囚犯被关进监狱，然后就从我们的视线中消失了。毫无疑问，暗箱操作广泛存在，但我们看不到。监狱里到底发生了什么？不要追问。现行的模型顽固地坚持一个依据可疑却不容置疑的假设，即高风险囚犯的囚禁时间越长，我们就越安全。如果有研究旨在推翻这种假设，那么这种研究很容易就被忽视了。

这就是现状。我们来看一下密歇根大学经济学教授迈克尔·穆勒–史密斯的再犯研究。在对得克萨斯州哈里斯县的260万份刑事案件的法庭记录进行研究后，他得出以下结论：在得克萨斯州哈里斯县，服刑时间越长，囚犯在获释后找不到工作的可能性就越大，因此他们就需

要更多的食品券和其他的公共援助，而且更可能会犯下更多的罪行。但是，要想利用这项研究让政策更加明智，社会更加公正，政客们将不得不考虑那个令他们畏惧的少数族裔群体的利益，而许多（也可能是大多数）选民更愿意忽视这一点。



“拦截—搜身”政策具有侵犯性，而且不公平，但在短时间内，该政策还将继续实行下去。这是因为警方正持续地从全球反恐行动中借鉴工具和技术，并将其用于打击地方犯罪上。例如，在圣迭哥，警察不仅要问被拦截人的身份，对他们进行搜身，有时，他们还会用iPad拍下被拦截者的照片，并将其发送到一个建立在云端的面部识别服务器，将照片与数据库中的罪犯和犯罪嫌疑人进行比对。根据《纽约时报》的一篇报道，圣迭哥警察局在2011~2015年间使用该面部识别程序对20600名被拦截者进行了面部识别，其中的大部分人还被采集了DNA信息。

由于面部识别技术的进步，更广泛的监测不久将成为现实。例如，波士顿警察局曾考虑过为户外音乐会设置监控摄像头以进行面部识别。这些数据将被上传至一个面部识别服务器，识别程序在1秒内能将一张脸与100万张脸进行比对。但最后，警方决定不这么做。在这一案例中，对隐私的保护胜过了效率。然而现实情况并非总是如此。

随着技术的进步，监控的程度和范围肯定会急剧增加。好消息是（如果你愿意称之为好消息的话），一旦安装在城市和乡镇的成千上万个监控摄像头拍下我们的面孔图像用于分析识别，警察就无须自己去辨认嫌犯了。毫无疑问，这项技术将有助于追踪嫌犯，波士顿市马拉松爆炸案的嫌犯就是靠这种技术抓住的。但是，这意味着我们都将遭遇一种数字形式的“拦截—搜身”——我们的面孔会被拍下并与数据库中已知的罪犯和恐怖分子相比对。

另外，监控的目标很可能会转变为发现、定位潜在的违法人员，这种定位不是在地图上划定一个社区或某个范围，而是直接指向具体个人。这些在反恐斗争中业已发展成熟的先发制人的策略，正是数学杀伤性武器的摇篮。

2009年，芝加哥警察局获得了来自美国国家司法研究所的200万美元拨款，用于开发犯罪预测模型。芝加哥警察局成功中标的理由是，他们声称只要有足够多的研究和数据，他们就能解释犯罪的传播模式。就像流行病的传染模式一样，这种模式是可以预测的，而且很有希望能够避免。

该项目的专家领头人是迈尔斯·沃尼克，他是伊利诺伊州理工学院医学影像研究中心的主任。几十年前，沃尼克曾通过分析数据帮助美国军方选择战场打击目标。此后，他转去从事医学领域的数据分析，包括探究老年痴呆症的病情发展情况。但和大多数数据科学家一样，他并不认为自己的专业技能与某个特定行业有联系。他的工作是发现模式——各种各样的模式。他在芝加哥开展的犯罪预测项目的研究重点是犯罪和罪犯的模式。

沃尼克团队早期也忙于锁定犯罪事件高发地区，就像PredPol的预测软件的开发者所做的那样。但是他们并没有止步于此，他们列出了一个最有可能犯下暴力罪行的人员的名单（人数约400人），并根据这些人犯下谋杀罪的可能性来进行排序。

名单上有一个22岁的高中辍学生，名叫罗伯特·麦克丹尼尔，在2013年的一个夏日，他打开家门，然后迎面撞到守在门外的一名警察。麦克丹尼尔后来对《芝加哥论坛报》说，他从未非法持有枪支，也从未犯下过暴力罪行。就像住在其所在的治安极差的社区奥斯丁里的大多数年轻人一样，麦克丹尼尔的确犯过一些小错误，并且认识很多进过监狱的人。他说，那个女警察告诉他，他现在正处在监控之下，要小心行事。

借助麦克丹尼尔在社交网络上留下的痕迹，警方将他归入了潜在重犯名单。他认识不少罪犯，而人们更可能和自己长时间待在一起的人的行为趋于一致，这不可否认。例如，脸书发现，在其平台上经常交流的用户更有可能点击同一条广告。从统计学上来说，“物以类聚，人以群分”是对的。

公平点儿说，芝加哥警察并没有直接逮捕像罗伯特·麦克丹尼尔这样的人，至少目前还没有。在这个案例中，警察的目标是拯救生命。如果400名最有可能犯下暴力罪行的人接到了警察的警告，他们中的一些人也许会在真正实施犯罪前三思。

但是，让我们从公平的角度思考一下麦克丹尼尔的案例。很不幸，他生长在一个贫穷又危险的社区。他的身边发生过很多罪行，他认识的很多人都有过犯罪记录。因为这些，而不是因为他自己的行动，他被认为是危险人物。现在，警方已经盯上了他。如果他表现不乖，比如购买了毒品，参与了酒吧打架，或者携带未注册的手枪，他会立刻受到法律的制裁，而且其遭受的处罚可能比大多数罪犯更严苛。毕竟，他已经被警告过了。

我要质疑的点是，使警察找到罗伯特·麦克丹尼尔的模型在目标设定上就是错误的。警察应该尝试在社区内建立关系，而不是试图简单地根除犯罪。这正是最初的“破窗理论”的主要主张之一。警察们要真正走进社区，和社区居民交谈，尝试帮助他们维护他们自己的社区标准。但是，这个目标往往被那些将逮捕更多罪犯等同于维护了社会治安的模型完全无视了。

好在，也并非处处都是如此。最近，我访问了新泽西州的卡姆登市，卡姆登市是2011年美国的“谋杀之都”。卡姆登市警察局于2012年重建，接受国家控制，其肩负双重使命：降低犯罪率和建立社区信任。如果建立信任被视作目标之一，那么逮捕就应该是最后而不是首选的手段。这项更具同理心的政策应该会让警察和民众之间的关系趋

于和缓，如此，像近年来我们看到的警察击毙年轻的黑人男子随后引发暴动的悲剧事件就应当会越来越少了。

然而，从数学的角度来看，信任是很难量化的，这对建模者来说是一个挑战。可悲的是，相较之下，计算逮捕数量、建立基于“人以群分”假设的模型要简单得多。而和罪犯生活在同一街区的无辜人士则因此遭受了极为不公的对待。由于贫困和犯罪记录之间的紧密联系，穷人将继续受困于这些数字法网难以逃脱，而未受影响的大多数则不愿意花时间考虑这个问题。

第六章
筛选

颅相学的偏见强化



几年前，范德堡大学一个名叫凯尔·贝姆的年轻人退学了。他患上了躁郁症，需要时间接受治疗。一年半之后，凯尔康复，去了另一所学校继续学习。大约就在那个时期，他从一个朋友那里获悉，有一个在克罗格公司做兼职工作的机会。这份兼职只不过是一个低薪的超市工作，他认为自己被录用肯定没问题。而且他的朋友正好要从这个岗位离职，还可以为他做担保。对于像凯尔这样的优秀学生，申请应该只是走个程序。

但是，凯尔没有接到面试邀请。他询问他的朋友原因，他朋友解释说，他被在申请这个职位时所做的人格测试“亮了红灯”。该人格测试是克罗格公司招聘筛选程序中的一个环节，这个程序是由总部位于波士顿郊区的人力资源管理公司克罗诺思（Kronos）开发的。凯尔跟他的父亲罗兰说了这件事。罗兰是一名律师，他问凯尔测试中都有什

么问题。凯尔说，那些题目很像他在医院做过的“大五人格”测试，后者就外向性、宜人性、严谨性、神经质和开放性五个方面为测试者打分。

起初，罗兰认为因为一份问卷而错过一份低薪工作似乎并不是什么大不了的事，他敦促儿子试试申请其他的招聘职位。但结果，凯尔每次都因同样的理由失败而归。他申请的公司都在使用同样的测试，他没能收到任何一份录用通知。罗兰·贝姆后来回忆道：“凯尔对我说：‘我的高考几乎满分，几年前我是范德堡大学的学生。如果连一份低薪的兼职工作都申请不到，那我是有多么糟糕？’我说：‘我觉得你没那么糟糕。’”

但是，罗兰·贝姆仍然感到很困惑。看起来，心理健康问卷好像在阻碍他的儿子就业。他决定研究一下这个问题。他很快就发现，大公司的招聘普遍采用了这种人格测试。而几乎没有任何人对公司的这种操作的合法性产生过质疑。他向我解释说，那些被“亮了红灯”的求职者很少知道他们之所以被拒绝是因为他们的心理测试结果不达标。即使他们知道，他们也不太可能聘请律师追究此事。

罗兰·贝姆陆续给7家公司发了通知，包括终点线（Finish Line）、家得宝、克罗格、劳氏、聪明宠物（PetSmart）、沃尔格林集团和百胜餐饮集团，他在通知中声明，他将提起集体诉讼，指控上述各公司在招聘过程中使用人格测试属于非法操作。

在写作本书的时候，案件正处于审理过程中。法庭争论的焦点大多集中在克罗诺思公司开发的人格检测是否可以被归为身体检查上，根据1990年的《美国残疾人法案》，公司在招聘时使用身体检查是非法的。如果有证据证明人格测试可被归类为身体检查的话，那么法院就必须针对涉案各公司是否应该为违反了1990年《美国残疾人法案》负法律责任做出裁决。

本章要探讨的是，在我们找工作的时候，自动系统会如何评判我们，以及它们评估的标准是什么。我们已经在之前的章节中看到，数学杀伤性武器严重影响了包括富人家庭子女和中产阶级家庭子女在内的准大学生的大学申请过程。与此同时，刑事司法系统中的数学杀伤性武器则将数百万人投进监狱，其中大多数囚犯都是穷人，根本没有机会上大学。这些群体面临着截然不同的难题，但他们也有共同之处：都需要一份工作。

以前，找工作往往取决于你认识谁。事实上，凯尔·贝姆去克罗格公司应聘超市兼职工作时也是按照这一传统路线进行的。他的朋友提醒过他在面试时要开放一点儿，并且向公司力荐了他。多年来，人们在杂货店、码头、银行或律师事务所的工作都是这样靠人与人的互相推荐得到的。接下来，应聘者通常需要参加面试，人力资源主管则通过面试来了解应聘者。不过，这种面试常常会变成人力资源主管的单一价值判断：这个人和我（或者和我的朋友）是一类人吗？结果是，与人力资源主管不是同类人的求职者有很大可能被拒绝，被拒绝的人中很多都是来自与主管不同种族、民族或有不同宗教信仰的求职者。女性也发现自己往往会被这场内幕游戏排除在外。

像克罗诺思这样的公司致力于人力资源工作的科学化，部分原因是为了让招聘过程更加公平。20世纪70年代，麻省理工学院的几位毕业生创立了克罗诺思公司，该公司的第一款产品是一种配备了微处理器的新型打卡钟，其能自动累计员工的总工作时间并上报给主管。这听起来可能有些老套，但它实际上是最早的追踪和优化劳动力工作效率的电子打卡设备（现在已经很发达了）。

随着技术的进步，克罗诺思公司开发了各种各样的职工管理软件工具，包括软件程序“劳动力管理套件”（Workforce Ready HR）。公司声称该产品将彻底取代“猜测式”招聘，并声明：“我们可以帮你筛

选、雇用和储备最有潜力的求职者，也就是那些符合公司需求、业绩更出色、对公司更忠诚的员工。”

克罗诺思公司是科技化招聘这一新兴产业的一分子。科技化招聘采用自动化程序，其中许多新近开发的招聘程序都包含人格测试，就像凯尔·贝姆做过的那项测试。一家名叫霍根测评的公司发布的数据显示，克罗诺思公司的年度营业额为5亿美元，并且仍在以每年10%~15%的速度增长。在美国职场中，目前约有60%~70%的求职者做过这样的人格测试，此数字比5年前上升了30%~40%。

这些招聘模型当然不可能包含应聘者将在公司有怎样的表现的信息，因为那是发生在未来的事，目前尚不可知。所以，像很多其他的大数据模型一样，招聘模型只能用替代变量预测员工的工作表现。我们在前面的章节已经了解到了，替代变量必定是不够精确的，而且往往隐含着不公平。其实，早在1971年，美国最高法院在格罗戈斯诉杜克能源公司案中就已经做出裁决，公司在招聘时采用智力测试具有歧视性，属于非法行为。有人可能会认为此案的判决很有警示作用。但恰恰相反，公司只是转向了使用智力测试的代替品，比如被“亮了红灯”的凯尔·贝姆所做的那项人格测试。

暂且不谈公平和合法问题。研究表明，人格测试并不能预测一个人的工作表现。为评估各种测试对工作表现的预测准确程度，美国艾奥瓦大学商学院的教授弗兰克·施密特分析了近100年的员工表现方面的数据。结果显示，人格测试的预测准确度很低，只相当于认知测试的1/3，远低于背景调查。但与此同时研究还表明，某些人格测试确实能帮助员工进一步了解自己，也能用于加强团队建设和成员沟通。求职者会在特定人格测试所创设的环境里明确思考该如何进行团结协作，而仅仅是要团结的这种意图就能够改善工作氛围。也就是说，如果测试的目标是找到积极乐观的员工，那么人格测试的确是一种有用的工具。

但是，人格测试并没有被用于此目的，而是被用作过滤器淘汰部分求职者。“人格测试的首要目标不是去寻找最好的员工，而是以尽可能低的成本排除尽可能多的人。”罗兰·贝姆说道。

你可能认为人格测试很容易被钻空子。如果你上网去做一个大五人格测试，你会发现题目好像很简单。有一个问题是：“你的情绪波动大吗？”回答“非常不符合”看起来是一个明智的选择。对另一个问题：“你容易愤怒吗？”明智的回答还是“不符合”，因为没什么公司想要聘用暴脾气的员工。

其实，通过这些问题筛选求职者的企业已经遭遇了麻烦。美国罗得岛州的监管部门发现，西维士医药公司涉嫌以非法手段筛选患有精神疾病的求职者。该公司的人格测试要求求职者对类似以下陈述回答“是”或“不是”：“你经常会因为别人做的事情而生气”，“结交好朋友没什么意义，他们总是让你失望”。不过，问题越复杂，钻空子就越难，招聘公司遭遇的麻烦也就越小。因此，如今公司使用的许多此类人格测试往往需要求职者做出一系列艰难的选择，让他们陷入“选是也不对，选不是也不对”的两难境地。

比如，麦当劳公司在招聘时就要求求职者在下列两项描述中挑选出更符合自己情况的一项：“当有很多事情要处理的时候，我会心情压抑”或“我有时候需要别人督促才会开始工作”。

《华尔街日报》邀请工业心理学家托马斯·查莫洛–普雷姆兹克来分析这类复杂的问题。针对上面那个问题，查莫洛–普雷姆兹克分析说，第一个选项与“神经质和尽责性”相关，第二个选项则与“低欲望和低动力”有关。换句话说，求职者要么就是在表明自己极易焦虑，要么就是在表明自己懒散成性。

克罗格公司的招聘问题相对简单一些：哪个形容词更能描述你的工作状态，“特立独行”还是“井井有条”？

查莫洛–普雷姆兹克对此分析说，选择“特立独行”表明此人“自我意识强、外向和有自恋倾向”，选择“井井有条”表明此人做事尽责，有较强的自控力。

注意，没有“以上两者都是”的选项。求职者必须要在不知道招聘模型将如何解读他的答案的情况下做出选择。而对于这些问题的解读有时会导向一些令人不快的结论。比如说，你在大多数幼儿园都会听到老师对班里的孩子说他们特立独行，这是一种褒奖，主要是为了增强孩子们的自信心。当然了，孩子们确实都是独特的。但是12年以后，如果那个已经长大了的孩子在为申请低薪工作而做的人格测试中选择了“特立独行”这个选项，招聘模型可能就会对这个答案“亮红灯”：谁想招聘一个自恋的员工?

人格测试的捍卫者指出，他们在测试中设置了很多问题，不会因为一个问题的答案不符合要求就刷掉一个求职者。但是，总归有某个特定的整体答案模式是被用来刷掉求职者的，而且我们不清楚那种答案模式到底是什么样的。我们没有被告知人格测试的目的是什么。整个评估的过程完全是不透明的。

更糟糕的是，一旦模型经过了技术专家的校准投入使用，此后模型收到的可用于修正的反馈数据就变得极少。这个模型与我们在第一章中讨论的用于体育运动的模型形成了强烈对比。大多数专业篮球队都会雇用数据极客，这些人负责通过分析球员的一系列参数建立表现模型，包括移动速度、垂直弹跳高度、罚球命中率以及大量其他的变量。当选拔运动新秀的时候，洛杉矶湖人队可能会拒绝来自杜克大学篮球队的一个表现亮眼的控球后卫，理由是根据表现模型，他的助攻数据不佳，而控卫必须是优秀的传球者。然而，在接下来的赛季里，他们可能会很沮丧地看到，那个不被看好的球员作为犹他爵士队的一员赢得了年度最佳新秀，并且他的助攻分数在联盟中位居前列。在这种情况下，湖人队就需要寻找模型中哪里出了问题。也许这个球员在

其大学校队主要负责得分，也许他在犹他爵士队学到了传球的诀窍。不管怎样，他们都可以用新的数据努力改善他们的模型。

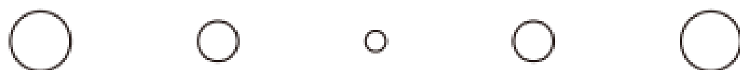
现在，设想一下凯尔·贝姆在被克罗格公司拒绝以后，在麦当劳找到了工作，并且成为那里的一名优秀员工。他在4个月内就被升职负责管理厨房，一年后就升为店长。那么克罗格公司的招聘人员中会有人回看人格测试的结果，调查他们是怎么犯了一个这么大的错误吗?

这是不可能的。这两种模型的区别在于：篮球队管理的是个体，每个球员的潜在身价可达数百万。篮球队的表现评估模型对其掌握竞争优势至关重要，而且篮球队特别需要反馈数据。如果没有持续输入的反馈数据，篮球队的模型就会失效。相反，雇用低薪员工的公司管理的是群体。它们用筛选机器取代人力资源专业人员，以降低成本。模型的目标是从众多求职者中筛选出更容易管理的那部分人。除非工作场所发生失控事件，比如大规模盗窃事件，或者生产效率急剧下跌等，否则公司没有理由去调整筛选模型。错过一些优秀的求职者完全是可以接受的损失。

此类公司可能对现状很满意，但受苦的则是自动筛选系统的受害者。你可能已经想到了，我会将公司人力资源部所采用的人格测试归为数学杀伤性武器。首先，它们被广泛采用并且影响巨大。尽管克罗诺思公司的人格测试漏洞百出，大多数公司还是会在招聘中使用它。在人格测试发明之前，雇主毫无疑问也带有偏见，但是不同公司的雇主偏见程度不一样。有些公司本可能为凯尔·贝姆这样的求职者打开一扇门，但现在，这种情况越来越少见了。从某种意义上来说，凯尔是幸运的。大多数求职者，尤其是应聘低薪岗位的求职者，常常被拒绝，但很少知道自己被拒绝的原因。而在偶然的机会下，凯尔的朋友得知了凯尔被拒的原因，然后告诉了凯尔。即便如此，如果不是因为凯尔的父亲是律师，有足够的时间和金钱应对大量的法律难题，起诉

使用克罗诺思公司开发的招聘软件的那些大公司也是几乎不可能的。低薪岗位的求职者很少有这样的家庭条件。(注)

最后，思考一下克罗诺思公司推出的人格测试造成的恶性循环。有某些心理健康问题的人被“亮了红灯”，他们因此没能找到一份正常的工作，过上正常的生活，这就使其进一步被社会孤立，而这正是《美国残疾人法案》要极力避免的情况。



谢天谢地，大多数求职者没有被自动筛选系统拦截，但是要从众多申请者中脱颖而出得到面试机会，仍是一个挑战。对于少数族裔和女性群体而言，他们面对的挑战要困难得多。

2001~2002年，简历自动阅读器还没有普及，芝加哥大学和麻省理工学院的研究人员制作了5000份虚假简历，应聘发布在《波士顿邮报》和《芝加哥论坛报》上的空缺职位，这些职位包括行政工作、客服和销售。每一份简历都包含了具体的种族信息。一半的简历申请人的名字被设计成白人的常见名字，如艾米丽·沃尔什、布兰登·贝克，另一半简历申请者的名字则看起来更像是非裔美国人，如拉奇莎·华盛顿、贾马尔·琼斯。研究者发现，白人名字的简历收到的回复比黑人名字的简历多50%。不过，第二个发现可能更加惊人：白人申请者中的履历优秀者得到的反馈比履历普通者要多；也就是说当面对的是白人申请者的简历时，招聘经理会给予简历的内容以更多的关注。但是，黑人申请者中的履历优秀者并没有比履历普通者得到更多的关注。显而易见，招聘市场依然带有很深的种族偏见。

避免此类偏见的理想方式就是盲选。长期以来由男性主导的管弦乐团，在20世纪70年代举行了一场著名的乐手试演会，乐手在演奏时需要隐藏在幕后。人脉关系和声誉突然间变得毫无意义，乐手的种族

或者毕业学校也变得不再重要，只有从幕后传来的音乐具有说服力。从此，女性乐手在管弦乐队的人数比例大幅跃升，现在已比当初增长了5倍，不过依然只占据乐团总人数的25%左右。

问题是很少有职业可以像乐团那样找到一种完全公平的选拔方式。幕后的乐手可以实实在在地演奏乐团的表演曲目，也就是他申请的这份工作的工作内容，无论是德沃夏克的大提琴协奏曲还是波萨诺伐舞曲的吉他演奏。而在除此以外的其他所有职业中，公司雇主不得不靠筛选简历寻找具有成功潜质的员工。

正如你料想的那样，人力资源部正依赖自动化系统筛选大量的简历。实际上，约72%的简历根本没有被招聘人员看到。电脑程序被用于挑选出具有雇主需要的技能和经历的简历，然后给这些筛选出来的简历评分，和空缺的职位进行匹配。筛选和匹配截止到哪个人的确仍是由人力资源部的人决定的，但是第一轮自动筛选排除的人越多，处理挑选出来的简历所需要的时间就越少。

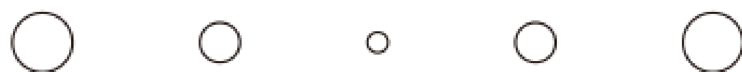
所以，求职者必须要精心制作自己的简历，要考虑简历筛选程序的偏好设定。举个例子，在简历中写入与空缺职位的工作内容相关的关键词是很必要的，这些关键词可以包含应聘的职位（如销售经理、首席财务官、软件构架师），语言技能（如普通话或者爪哇语），以及获奖情况（如优秀毕业生、鹰级童子军）。

懂行情的人知道简历筛选程序喜欢什么不喜欢什么。举个例子，图片毫无意义。大多数简历筛选程序尚不会处理图片。求职平台Resunate.com的共同创始人莫娜·阿卜杜勒–哈利姆说，花哨的字体不仅没什么用处，还会加大机器的识别难度。她说，安全的简历都采用普通字体，如Ariel和Courier字体。另外，不要使用箭头之类的符号，它们也只会加大自动系统正确解析信息的难度。

这些程序和大学录取程序一样，其所造成的结果就是那些拥有充足的金钱和资源来准备简历的人脱颖而出。没有精心准备简历的人也许永远不会知道他们投递的简历已经早早被机器排除了。这个案例再一次生动地展示了有钱人和消息灵通的人能占据多大的优势，而穷人被判出局的可能性又有多高。

公平地说，筛选简历这件事总是会存在这样或那样的偏见。上一代求职者中的消息灵通人士会非常细心地把简历内容写得明晰连贯，用IBM（美国国际商用机器公司）出品的高端电脑输入信息，再用优质打印纸打印出来。因为这样的简历获得招聘者认可的可能性更高。手写的简历，或者蹭上油印机器黑墨的简历通常都会被丢进垃圾桶。所以，从这个意义上来说，机会的不公平并不是什么新鲜事，只是现在不公平有了一个新的化身——自动化筛选程序。

这些自动化筛选程序的不公平操作不仅仅局限于筛选简历。我们的生计也越来越依赖于我们顺应机器的能力。最明显的例子就是谷歌。不管是连锁酒店还是汽车修理厂，一家企业的成功在很大程度上取决于它是否能够出现在搜索引擎的第一页。现如今，个人也面临着同样的挑战，不管是进入一家公司，向上晋升，还是只是逃过裁员浪潮，这一切的关键就在于了解筛选机器的判定标准是什么。可是，在这个被吹捧为公平、科学和民主的数字化世界里，仍然只有内部人士才能找到那条通往成功的路径。



圣乔治医学院位于伦敦南部的图庭区，20世纪70年代，该医学院招聘办公室发现了一个机会。招聘办每年发布约150个职位空缺，每一个职位都会收到超过12份申请。梳理所有这些申请工作量巨大，需要很多人手。而且，因为每一个筛选者都有不同的想法和偏好，整个筛

选过程多少显得有点儿随意。那么，是否有可能通过编写一个电脑程序来筛选申请，把申请数量缩减到一个更可控的范围内呢？

当时，大型机构和公司已经在利用电脑处理这样的工作了，比如美国国防部和IBM。但是，在20世纪70年代末，一所医学院说要开发一套自动化评估系统则可以说是一个非常大胆的实验，就像苹果公司要发布其第一台个人电脑一样。

然而事实证明，圣乔治医学院的实验彻底失败了。圣乔治医学院不仅在数学建模方面不够成熟，而且在不知情的情况下成了数学杀伤性武器的“开拓先锋”。

和很多数学杀伤性武器一样，当管理层为模型同时设定了两个目标时，问题就出现了。第一个目标是提高效率，让机器处理大量烦琐、重复的工作，自动从2000份申请中筛选出500份，然后再由人工负责面试工作。第二个目标是公平，保证模型不受管理层的情绪或者偏见的影响，以及不受上院议员或者内阁大臣的人情关系的影响。因此，在第一轮自动筛选中，针对每一个申请者的评判标准将是一样的。

那么评判标准是什么呢？这个问题似乎很简单。圣乔治医学院保留着大量往年的筛选记录，其目标就是教程序系统复制人类以往一直遵循的评判标准。我相信你们已经猜到了，这些数据的输入正是问题所在。电脑从人这里学到了如何区别对待不同的申请，而且电脑进行区别对待的效率惊人。

公平地说，并不是圣乔治医学院所有的训练数据都存在明显的种族歧视。许多写着外国名字或者外国地址的申请，可以很明显地看出申请者还没有掌握英语，而筛选机器只是简单拒绝这些申请，而不会考虑其中优秀的申请者可以很快学会英语的可能性。（毕竟学校必须筛掉3/4的申请，而从外语使用者入手看起来是个不错的选择。）

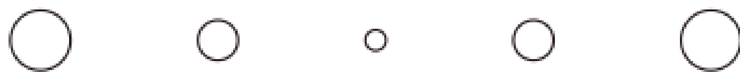
现在的问题是，虽然以往圣乔治医学院的工作人员可以筛掉含有语法错误和拼写错误的申请，但这对计算机来说有点儿太难了。但是，训练数据让它在之前被筛掉的申请与申请者的出生地和姓氏之间建立了联系。所以，来自非洲、巴基斯坦和英国移民聚居地等地区的人得分更低，更可能不被给予面试机会，而这些人中的大部分不是白人。以往的招聘人员也会拒绝女性求职者，他们理所当然地认为，家庭责任和抚养孩子很可能会影响她们的工作表现，而这些当然也被机器学会了。

1988年，英国政府的种族平等委员会发现，圣乔治医学院的招聘政策涉嫌种族和性别歧视。该委员会的调查数据显示，每年2000个申请者中有60个申请者无法进入面试环节，仅仅是因为他们的种族、民族或者性别。

为了消除偏见，圣乔治医学院以及其他行业的统计学家可以设计一个数字版的盲选系统，排除地理位置、性别、种族或者姓名等变量，只关注和医学教育相关的数据。关键是去分析每一个申请者能为学校带来什么技术，而不是将一个申请者放在与其相类似的其他申请者中做比较。而且，圣乔治医学院本可以开创性地解决女性申请者和外籍申请者面临的难题。《英国医学期刊》的报告以及种族平等委员会都提出过相关的办法。如果优秀申请者存在语言和育儿上的困难和负担，合适的处理方法不是直接拒绝他们，而是帮助他们解决问题，如给他们提供英文课程或者校内托儿所。

在接下来的章节中我还会再讲到这一点，即我们已经多次看到，数学模型可以通过数据筛选锁定那些很可能正在遭遇人生难题的人，不管人生难题是犯罪、贫困方面的，还是教育或其他方面的。利用数学模型拒绝和惩罚他们，还是给他们提供需要的资源去帮助他们，完全取决于社会。规模和效率使得数学杀伤性武器更具破坏力，但我们

也可以利用模型的规模和效率帮助这些人，如何选择完全取决于我们为模型设定的目标。



这一章到现在，我们一直研究的是筛选求职者的模型。大多数公司利用数学杀伤性武器降低行政开支，减少用人不善（或者需要更多培训投入）的风险。简单来说，过滤器的目标就是节省开支。

人力资源部门当然也非常希望选择一种能节省开支的招聘办法。公司运营的主要成本之一就是员工流动，或员工流失。根据美国进步研究中心的调查数据，一家公司找人接替一个年薪5万美元的员工的成本约为1万美元，也就是这位员工年薪的20%；接替一个高层管理者的成本还要翻上几倍——两年的高管年薪。

所以，自然会有很多招聘模型致力于计算求职者的忠诚度。云技术人才管理软件解决方案的全球领导者Cornerstone OnDemand旗下的公司Evolv帮助施乐公司（Xerox）的客服接待中心寻找潜在的合格员工，客服接待中心需要招聘4000多位职员。Evolv的员工流失模型考虑了一些相关变量，其中包括历次工作的平均服务时间，这一点你可能已经料想到了。但是，Evolv还发现了一些更有趣的联系：被系统归类为具有“创造性”的求职者往往更具忠诚度，而那些“好奇心”得分高的求职者则更可能跳槽。

但是，问题最大的联系再次涉及地理位置因素。该公司发现，家庭住所离公司更远的求职者更有可能辞职。这可以理解，长距离通勤是一种痛苦。但是施乐公司的管理层还注意到一个联系：许多苦于长距离通勤的人往往来自贫困街区。所以，施乐公司极为高尚地从模型中去掉了与工作表现高度相关的员工流失数据。施乐公司为了公平牺牲了一点儿效率。

员工流失分析是为了防止出现用人不善的情况，而人力资源部门更重要的工作是寻找未来的职场之星，也就是寻找那些具备能够改变公司命运的智慧、创造性和驱动力的人。大型企业迫切需求那些具有创造力和团队合作精神的求职者。所以，建模者面临的挑战是在大数据的广阔世界里准确描述出和创造性、社交技能相关的那一小部分信息。

申请者简历所包含的零星信息当然不够。简历中列举的大多数信息，如名牌大学、获奖情况以及相关技能，都只是高质量工作表现的极为不精确的替代变量。毫无疑问，技艺超凡和名牌大学学位两者之间有关联，但是这种联系并不是绝对的。很多软件开发人才并不是名牌大学毕业生，比如那些高中生黑客。而且，申请者在简历中往往夸大其词，还有可能撒谎。系统通过快速扫描领英或者脸书等平台可以了解更多有关求职者的信息，还能找到求职者的部分朋友和同事。但是，依然很难根据这些数据预测某个工程师是否是位于帕洛阿尔托或沃斯堡市的目前只有几名成员的初创咨询公司的绝佳人选。找到一个能胜任这种角色的人需要更全面的数据和更高水平的模型。

这一领域的模型开创者是位于旧金山的创业公司吉利德（Gild）。除了求职者的母校或者简历，吉利德公司还梳理了数百万个求职网站，分析每个求职者的所谓的“社会数据”。吉利德公司为其客户（大多数是科技公司）建立求职者的个人档案，保证客户及时了解哪些求职者又增加了新技能。吉利德公司声称，其开发的个人档案甚至可以预测其他公司中的一个明星员工什么时候可能想要换工作，并及时提醒客户公司在恰当的时机发出邀请。但是，吉利德公司的模型在做量化处理的同时，也为每一个职员赋予了“社会资本”。这个人在程序员同行中的重要性有多高？他会分享并贡献自己的代码吗？比如说，一个巴西的程序员，我们姑且叫他佩德罗，他住在圣保罗州，从每天晚上吃完饭到凌晨1点，他都会和地球另一端的同行交流，尝试解决GitHub或Stack Overflow网站上的云计算问题或者头脑风暴某个游

戏程序的算法。吉利德公司的模型会试图测量佩德罗对编程技术的热情（然后赋予他一个高分）和他与同行的互动程度，也会评估佩德罗的社交圈的技能水平和社会地位。有大批追随者的人更有价值。如果佩德罗的主要联系人里刚好有谷歌合伙人谢尔盖·布林或者虚拟实境头盔欧酷拉的创始人帕尔默·拉奇，那么佩德罗的社交得分毫无疑问将一飞冲天。

但是，吉利德公司所开发的模型以及类似的模型很少能从大数据中收集到像佩德罗这样明确的信号。所以它们采取广撒网的战略，搜索所有能找到的和职场名人有关的数据。借助公司数据库中的600多万个编码器的强大力量，它可以找到各种各样的模式。吉利德公司的首席研究员薇薇恩·明在接受《大西洋月刊》的采访时说，吉利德公司发现一群技术天才会经常性地访问一个日本漫画网站。如果佩德罗也经常浏览这个漫画网站，那么这一数据虽然不会直接给他贴上职场明星的标签，但的确会提高他的分数。

这对佩德罗是有意义的。但是，有些职场人员在线下做的事情即使是最高明的程序也推断不出来，至少在今天还做不到。比如说，他们可能在照顾孩子，或者出席一个读书小组。如果某个潜在的优秀员工没有每天晚上花6个小时讨论日本漫画，那么这不应该成为他的不利条件。正如科技领域仍然是由男性主导的一样，登录日本漫画网站的人也主要是男性，而且因为这些网站包含色情成分，所以科技行业中的大部分女性员工很可能对此完全没有兴趣，而关键在于，这不应该成为她们求职时的减分项。

尽管有以上这些问题，但吉利德公司只是该行业的众多竞争者之一，没有作为全球巨头的影响力，也没有资格将其模型设定为通行的行业标准。比起我们看到的一些更恐怖的模型，如让一个个贫困家庭债台高筑的掠夺式广告，把求职者拒之门外的人格测试等，吉利德公司的预测模型已经算是仁慈的了。即便吉利德公司的预测模型对我们

的有利面比有害面多，且这种模型无疑是不公平的：一些极有潜能的人才可能因为各种各样的因素而被模型理所当然地忽略了；但我仍然认为，这种程度的人才流失不至于到达数学杀伤性武器那样的危害程度。

还需要指出的是，这些用于招聘和人才储备的模型在不断进步。数据世界仍在不断扩张，我们每一个人都会在生活中制造大量的新数据。这些数据都将被提供给我们将来的老板，而他们将通过这些数据了解我们。

那么，老板对我们的看法会被检测吗？还是只被用来维护现状和强化偏见？每次看到某些公司使用数据的那种草率、自私的方式，我常常会想起风靡于19世纪的一种伪科学，颅相学。颅相学家用手指摸病人的头盖骨寻找肿块和凹陷。他们声称，每一个肿块和凹陷都和存在于人脑27个区域里的人格特质有关。颅相学家通过摸头盖骨得出的结论往往和他的观察相一致。如果一个病人极端焦虑，或者饮酒过度，颅相学家通常会在病人的头盖骨中发现与这些问题相对应的肿块和凹陷，这反过来又强化了人们对颅相学的信任。

颅相学也是一种模型，依靠伪科学建立权威，风行了几十年。现如今的大数据似乎也落入了同样的窠臼。对凯尔·贝姆“亮红灯”的人格测试模型、圣乔治医学院筛掉外国医学院优秀申请者的招聘模型都将人们拒之门外，而这些模型的“科学性”之所在只不过是一些未被证实的假设罢了。

---

1. 没错，许多中产阶级的准大学生也会在暑假做一两份低薪工作。但是，如果他们也遭遇了不好的求职经历，或者遭到一个武断的数学杀伤性武器的误判，那么这种情况只会强化如下观点：他们应该在学校学习，而不应该出来做这种糟糕的工作。

第七章
反馈

辛普森悖论的噪声



美国大公司的职员最近创造了一个新的动词：关开门（clopening），指的是一个职员工作到很晚，关闭门店或者咖啡店，几小时后又在天亮之前回来开门。站在公司运营的立场，安排同一个职员关门开门确实有一定道理，但是，它带来的结果是严重疲惫的员工，和逼人发疯的工作时间表。

极度不规律的工作时间安排越来越常见，首当其冲的是星巴克、麦当劳和沃尔玛等企业的低薪职工。对工作安排调整的不及时通知使得这一问题更加严重。许多职员提前一两天才接到通知要在周三上夜班或者在周五高峰时段值班。他们的生活因此变得一团糟，照顾子女的计划也颇受影响，吃饭和睡觉都得抢时间。

这种不规律的工作时间安排是数字经济带来的。在上一章，我们看到数学杀伤性武器是如何筛选求职者的，它们排斥某些人，漠视很多人。我们也看到，筛选模型是如何从过去的记录中学会了不公平的运作方式，将有害偏见编入模型之中的。在这一章，我们将会看到，在数字经济的大环境下，追求效率的数学杀伤性武器将职员当成了机器上的齿轮。“关开门”只是这种趋势的表现之一，随着监控设备在工作场所的普及，在数字经济因此而进一步发展壮大的同时，其带来的副作用也将进一步恶化。

在公司尚未开始采用大数据技术的很长时间里，工作时间安排根本不是什么科学。想象一下，一家个体五金店，店员每天的上班时间为早上9点到下午5点，每周工作6天。某一年，老板的女儿去上大学了。暑假回家后，她得以用一种新的眼光观察她家的五金店。她注意到，每周二的早上都没有顾客上门。这段时间，雇员一直拿着手机上网，无人打扰。而在每个周六，来店的顾客都会抱怨排队的时间太长。这显然不利于五金店的盈利。

这些观察提供了有价值的数据，老板女儿帮助她的父母调整了五金店的运作方式。首先，他们在每周二上午关闭五金店，然后在每周六的高峰时段加雇一个兼职人员来帮忙。如此，原来呆板、僵化的运营方式就变得更加智能了。

在大数据时代，负责分析的不是大学新生，而是许多博士，辅助以功能强大的计算机。如今，企业可以通过分析顾客流量，精确计算出每一天的每一小时它们需要多少人力。当然，该分析的目的是尽可能地降低成本，其有助于企业维持最低限度的人力，同时确保业务繁忙时段有充足的人力支援。

你可能会认为业务的繁忙和空闲模式每周都会重复，公司只需要对员工固定的工作安排做一些调整，就像我们假设的那家个体五金店所做的那样就可以了。但是，新的人力安排软件为公司提供了丰富得

多的选择。它们能够处理不断更新的数据流，从天气到路人的行为模型皆可纳入考量。举例来说，某个雨天的下午，公园里的人可能会因为想要避雨跑进咖啡店。此时，咖啡店就需要更多的人手，至少在下雨的一两个小时里需要。每周五晚上，如果有高中橄榄球比赛，那么大街上就会出现更多的路人，但是仅限于球赛前和球赛后的时段，而不包括比赛进行中的那个时间段。推特的流量分析预测，今年的黑色星期五去商店抢购的人数将比去年多26%以上。情况随时都在变化，企业必须合理配置人力以匹配不断波动的需求，否则就意味着浪费金钱。

通过调整人力安排省下的钱自然直接进入了老板的口袋。此前，在低效率的经营模式下，职员不仅有可预测的工作时间，而且还有固定的闲暇时间。你可以说他们受益于低效率的人力安排：有些人可以在工作时间读书，甚至做研究。而现在，在借助电脑程序进行的人力配置下，职员每一分钟都很忙。程序可能会随时给员工分配临时的工作安排，包括周五晚上负责下班关门、周六一大早负责开门。

2014年，《纽约时报》报道了一个饱受工作时间安排不规律困扰的单亲妈妈，她叫简奈特·纳薇欧。她一边努力读大学，一边在星巴克担任咖啡师，与此同时还得照顾她四岁的孩子。但她的值班时间不断改变，偶尔还必须负责关开店，这令她的生活陷入混乱。她无法为她的孩子安排固定的日间托儿服务，因而不得不暂停学业。她的时间只能用来上班。纳薇欧的情况在美国相当普遍。美国政府的统计数据显示，2/3的餐饮业劳工和超过一半的零售业劳工如果工作安排临时有变，其被提前通知的时间最多不超过一周，通常只会提前一两天，他们因此只能仓促地调整用车安排或托儿服务的时间。

在《纽约时报》的这篇文章发表后的几周内，被点名的大公司宣布将调整它们的人力调度方式。这篇报道令这些雇主尴尬不已，它们承诺，它们将在模型中加入一个限制条件，也就是不再要求同一名员

工负责关开店，接受效率稍差一点儿的人力安排。因为品牌形象更仰赖公平对待员工，星巴克决定多做一些，其宣布将调整其人力调度软件，减轻公司13万名咖啡师一直以来所承受的噩梦般的值班时间频繁变更的困扰。星巴克承诺，将至少提前一周公布员工的值班安排。

但是，《纽约时报》一年后的追踪报道显示，星巴克并未信守承诺，甚至连杜绝同一员工关开店的安排都没有做到。根本问题在于，维持最低人力成本已成为企业文化的一部分。在许多公司，管理层的薪酬都取决于他们的人力运用效率，以员工每工作一小时为公司创造多少营收为衡量标准。人力调度软件协助管理层提高这些数字以及他们的薪酬。即使公司高层要求前线管理者对员工的安排放宽一些，前线管理者也往往会抗拒，因为这违反他们一直以来被灌输的观念。星巴克一名员工表示，在星巴克，如果某店的店长在员工安排方面的投入超出了其被规定的“人力预算”，系统就会将此情况通知该区的经理，而该店店长很可能因此留下负面的书面记录。为避免这种情况，店长很可能会选择继续临时调整员工的值班时间，即使这会违反公司提前一周公布值班安排的承诺。

说到底，像星巴克这种上市公司的商业模型，都是以提高盈利为目的的。这一目的同时反映在了它们的企业文化和激励措施上，也对它们的运营管理软件产生了越来越大的影响。（而如果这种软件容许微调，那么所做的调整也更可能是出于提高盈利的目的而非其他。）

这种人力调度技术主要源自“运筹学”这门功能强大的应用数学。数百年来，数学家运用“运筹学”的基本技术，帮助了农民拟订种植计划，也帮助了土木工程师规划公路路线，提高人员和货物的流通效率。但这门技术直到“二战”时才真正发展壮大起来，当时美国和英国军方招募了多个数学家团队，以协助军方优化资源利用。盟军计算了各种形式的“交换率”，考虑在每个作战方案中盟军为了摧毁敌人的资源耗用了自己的多少资源。在1945年3月至8月的“饥饿作战”期间，第

21轰炸机司令部奉命摧毁日本商船，阻止食物和其他货品安全抵达日本控制的港口。盟军的运筹小队致力于研究如何以最少的轰炸机摧毁最多的日本商船，最终找到了能够超过40:1的“交换率”的作战方案：在轰炸行动中以仅损失15架飞机的代价摧毁了606艘日本船只。这次作战行动被视为一次非常高效的行动，而运筹小队对此有着显著的贡献。

“二战”之后，大企业（以及美国国防部）投入了大量资源到运筹学研究上。而由此产生的物流学从根本上改变了我们生产商品和将它们送到市场上的方式。

到20世纪60年代，日本汽车厂商又有了重大突破，其设计出了“准时制”（Just in Time）制造系统。在“准时制”系统中，汽车组装厂不会在仓库里事先储存大量的汽车零组件，然后让其中的大多数处于闲置状态，而是会根据需要实时向供货商订购零组件。丰田汽车和本田汽车都建立了复杂的供应链，实现了随传随到的零组件供应系统。汽车业仿佛成了一个完整的有机体，有了自身的体内平衡控制系统（homeostatic control systems）。

“准时制”系统非常高效，很快便在全球普及。许多地方的公司非常快速地建立起了准时制供应链。这些模型也成了支撑亚马逊、联邦快递和优比速快递等公司的商业运作的数学基础。

人力调度软件可被视为“准时制”经济的延伸，只不过，必须“随叫随到”的不再是割草机刀片或手机屏幕，而是人——而且通常是迫切需要金钱的人。由于这些人迫切需要金钱，其雇主似乎因此就被赋予了扭曲他们的正常生活，令其配合数学模型要求的权利。

我必须补充一点：企业也会采取一些措施避免员工的生活变得太过痛苦。如果员工受不了折磨而辞职，找到替代人选需要花多少钱，这一点，雇主们全都了然于胸。这些数字也在模型所搜集数据的范围

内。如上一章所述，企业另有一些模型用以降低员工流动率，因为员工的频繁流动会损害企业的盈利和效率。

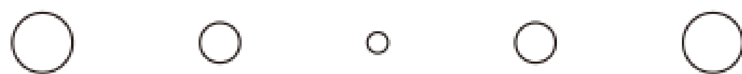
从员工的角度来说，问题出在低薪劳工供给过剩上。人们迫切需要工作，这也是许多人坚持不放弃一份时薪仅8美元的工作的原因。劳动力供给过剩，加上有影响力的工会极为少见，劳工几乎毫无议价能力可言。这意味着，大型零售企业和餐饮公司的雇主几乎可以随意扭曲员工的生活，令其配合越发荒谬的工作时间安排，而不会因此受到员工流动率过高的困扰。这些公司将赚到更多的钱，但其员工的生活则越来越恶劣。而因为这种人力调度优化程序无所不在，劳工深知换工作也不大可能改善自己的境况。总而言之，拜上述种种因素所赐，企业得到了可以轻易控制的劳动力。

因此我认为，人力调度软件是一种更为可怕的数学杀伤性武器。如前所述，这种模型的应用范围巨大，而且剥削的对象是本已在艰难谋生的人。另外，这种模型是完全不透明的。劳工往往完全不知道自己将在何时被叫去上班，因为召唤他们的是一个武断专制的程序。

人力调度软件也制造出了一种恶性循环。想想前文提到的单亲妈妈纳薇欧。因为工作时间非常不稳定，她不可能再回到学校念完大学，而这就损害了她的就业前途，令她只能一直处在供大于求的低薪劳工行列当中。因为工作时间又长又不规律，劳工也很难被组织起来，为争取较好的工作条件而抗争。相反，他们会变得格外焦虑，而且睡眠不足，后者令人更容易出现剧烈的情绪波动，而美国约有13%的高速公路交通事故是由睡眠不足引起的。更糟糕的是，因为这种软件是设计来为企业省钱的，其往往会将员工的每周工作时间限制在30小时之内，这样，公司就不必为员工提供医疗保险。而因为工作时间非常不稳定，这些劳工多数无法挪出时间做第二份工作。这一切看起来就好像设计这种软件就是为了惩罚低薪劳工，令他们无法出头一般。

这种软件也导致美国一大部分的儿童在成长过程中无法拥有一个正常的家庭环境。这些小孩只会在吃早餐时看到自己的妈妈睡眼惺忪，或是没吃晚餐就匆忙出门，又或者与外婆争论谁可以在周日早上照顾他们。这种混乱的生活对小孩的成长有着深刻的影响。美国智库经济政策研究所的一项研究指出："如果父母的工作时间不规律，又或者常在标准的日间工作时间之外外出工作，他们的孩子（儿童或青少年）的认知和行为表现就可能会低于平均水平。"当孩子在学校表现不好或出现各种行为问题时，父母可能只会责怪自己，但真正的罪魁祸首其实往往是导致这些家长只能去做工作时间不固定的工作的贫困处境，以及那些进一步压榨这些贫穷家庭的人力调度模型。

一如许多其他的数学杀伤性武器，问题的根源在于模型创建者选择了什么目标。这些模型追求的是效率和盈利的最大化，而不是正义或"团队"的福祉。当然，这是资本主义的本质。对企业来说，盈利有如氧气，是维持其生命力的必要条件。站在他们的角度，可以省钱而不省是极其愚蠢的，甚至是违反自然规律的。这就是为什么社会需要一些与此对抗的力量，例如利用媒体报道有力地凸显过度追求效率的恶行，令相关公司感到羞愧，借此促使它们去做正确的事。而如果这些公司未能充分纠正错误（就像星巴克那样），则媒体要做的就是反复揭露这些问题直到改变发生。此外，我们也需要监管部门有效地约束企业，强大的工会组织劳工、放大劳工的诉求，以及政治家立法遏制企业的恶劣行为。在《纽约时报》2014年刊出相关报道后，美国国会中的民主党人迅速拟定了约束人力调度软件的法案。但是，因为在国会占据多数的共和党人强烈反对政府规管企业，该法案最终没能通过。



2008年，正当经济大萧条迫近之时，一家位于旧金山的公司Cataphora开发出了一个软件系统，该软件可以根据若干变量评价技术

工作者的表现，包括他们在生产创意方面的表现。这不是一项简单的工作，因为软件程序难以在创意和一串简单词语之间做甄别。稍微思考一下你就会知道，二者的区别往往体现在语境上。昨天的创意（地球是圆的，在社交网络分享照片等）已经变成了今天的现实。我们每一个人都多少能明白一个创意在什么时候变成了既定的事实以及在什么时候被证明了是一文不值的（尽管我们常常排斥这种判断）。但是，即使是最高端的人工智能在这方面也束手无策。所以，Cataphora公司的系统需要向人类寻求指导。

Cataphora公司的软件所做的主要工作是探查企业邮件、短信的往来。该软件的指导假设是，越好的创意越容易在互联网上得到广泛的传播。如果人们复制了一些文字并在社交网络上分享了它们，那么这些文字可能就是创意，软件可以量化这些创意的传播程度。

但是，实际情况比较复杂。创意不仅仅是在社交网站上被分享的一段段文字。举例来说，笑话一般也会流传甚广，而软件系统很可能将其与创意混为一谈。八卦也传播得很快。但是，笑话和八卦遵循一定的模式，程序可以通过学习这些模式将其筛除出去。久而久之，系统就能逐渐辨认出一段段更有可能代表创意的文字。软件系统通过网络追踪创意，计算它们被复制的次数，测量它们的分布，找出它们的来源。

很快，公司内部的角色分工就变得清晰明了了。系统得出结论，一部分人是创意生成者。公司在其员工图表上用圆圈圈出创意生成者，其中提出过很多创意的员工，其圆圈会更大、更黑。另一部分人则充当着连接器，就像分布式网络中的神经元，他们负责传递信息。最高效的连接器会将一段段文字迅速传播开来。该系统同样用深色将后者标注出来。

不管该模型是否能有效地评估创意流，其假设本身没有危害。利用这种分析来辨认人们的想法，为他们匹配最适合的同事和合作者是

件有意义的事情。IBM和微软公司就在公司内部采用了类似的程序进行同样的操作。这就像是一个约会对象配对算法（并且无疑也会产生类似的各种各样的结果）。另外，大数据也被用来研究客户服务中心的业绩。

几年前，麻省理工学院的研究员分析了美国银行客户服务中心的职员的行为，目的是调查某些组的业绩比其他组更好的原因。他们让中心的每一个职员佩戴一个社交测量胸牌。胸牌内的电子记录仪会追踪职员的位置，每16毫秒测量一次他们的声调和姿态，记录职员之间什么时候对视，以及每个人说话、倾听和被打断的频率。客服中心4组职员总共80人，每人都需要佩戴胸牌6周。

这些职员受到的管制特别严格。公司不允许员工闲谈，因为客户服务中心的职员应该尽可能多地接听电话，解决客户的问题。公司还要求，休息时间员工不能结伴出去，要一个一个去。

研究员惊讶地发现，业绩最好、工作最高效的职员同时也是最擅长社交的职员。这些职员忽视规定，比其他职员的闲谈时间更长。据此，公司鼓励所有客服中心职员多参与社交，该部门的业绩随之飙升。

但是，追踪职员行为的数据研究也可以用来裁减劳工。2008年金融危机期间，技术行业的人力资源部门采用了一种新的视角看待Cataphora模型的结果图表。他们发现，某些员工会被大大的黑色圆圈标注出来，而另一些则被用较小较浅的圆圈标示。如果不得不裁员（很多公司已经裁员了），先开除后者显然是一种理智的选择。

但是被开除的职员是真的没有价值吗？这里我们要再次提及数字颅相学。我们现在看到的是，如果一个系统认定一个雇员为低创意者或者低连接者，那么这一系统给出的结论就变成了事实——这就是这位雇员的评分。

但也许某些职员就是例外。被浅色圆圈代表的职工也许也曾提出过很多绝妙的想法，只是没有把它们分享在网络上，比如，某个职员可能在午餐时间提出了一个绝佳的点子，或者以开玩笑的方式打破了办公室的紧张气氛，这让同事们都很喜欢他，这一点在职场上非常有价值。但是，计算机系统无法将这些抽象的软技能数字化。软技能数据没有被收集，而且即便收集了也难以赋值，所以往往会被模型遗漏。

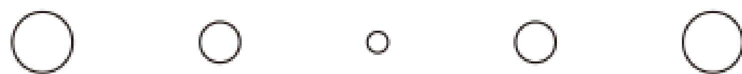
所以，该系统只是找出了表面上的失败者。他们中有许多人因此在2008年经济大萧条期间丢掉了工作。仅这一点就已经是不公平了。但是，更糟糕的是，如Cataphora模型这样的系统收到的反馈数据也是最少的。有人被鉴定为失败者，然后被开除，也许他之后又找到了其他的工作，发明了很多专利，但这些数据往往不在系统的收集范围之内。系统根本不知道自己做了多少错误的判断。

这是一个大问题，因为科学家需要错误反馈来做取证分析，查明系统哪儿出错了，什么变量被误读了，什么数据被忽略了。而系统则在此过程中得以学习并进化。但是，如我们所见，大量的数学杀伤性武器，从再犯模型到教师评估模型，只不过是在打造它们自己所定义的现实。管理层假定教师评估模型是贴合现实情况的，因此认为它非常有用，该模型使得考核教师的教育成果变得更加容易。管理层可以据此开除职员，削减开支，把责任推卸给一个客观的评分数字，而不管这一数字是否真正符合现实情况。

Cataphora公司的规模很小，开发员工评估模型只是该公司的副业，其最主要的业务是查明欺诈或公司内线交易的模式。Cataphora公司于2012年宣布破产，公司将员工评估软件卖给了一个名叫Chenope的初创公司。但不论如何，此类软件的确存在着成为真正的数学杀伤性武器的潜力。这些软件会错误地解读员工，惩罚失败者，且根本无法证明它们的评分和员工的工作能力存在关联。

这种软件标志着数学杀伤性武器在新领域的崛起。多年来，似乎只有工厂的工人和服务业的雇员会被模型化、最大化利用，而律师、化学工程师等高智商群体则能够避免数学杀伤性武器的伤害，至少在工作当中如此。而实际上，Cataphora公司的软件早就预示着情况将发生改变。现在，科技行业有很多公司正通过研究白领职员的交流模式来最大化地挖掘并利用他们的潜在价值。谷歌、脸书、亚马逊和IBM等科技巨头尤其热衷于使用最优模型。

不过，至少到目前为止，多样性仍在被强调。这至少为我们保留了一个希望，即被某一模型给出差评的员工有可能被另一模型给出高分评价。但最终，一个统一的行业标准将会出现，那时，我们都将陷入一个无望之地。



1983年，里根总统时代，美国教育当局发表过一篇著名的名为《国家处在危险中》（*A Nation at Risk*）的报告，报告指出，美国教育正被“一股平庸的浪潮”侵蚀，它威胁着整个“国家和民族的未来”。

报告还写道，倘若“一股怀有敌意的外国势力”试图将今天存在着的平庸的教育实绩强加在美国身上，“我们会把这视作战争行动”。

最明显的信号就是学生高考分数直线下降。1963~1980年，高考阅读平均成绩下降了50分，数学平均成绩下降了40分。在全球经济竞争的背景下，我们的竞争力仰赖我们的技术，但技术进步的基础似乎越来越脆弱。

谁应该为这种糟糕的状况负责？报告把责任甩给了教师。报告呼吁国家采取行动，让学生参加前测试，然后根据学生的考试成绩开除不合格的教师。就像我们在本书引文中所看到的那样，大批教师因此丢掉了工作。华盛顿学区教师萨拉·韦索基就是这项行动的受害者，她

因学生考分极低而被学校开除。我讲述这件事的目的是展示数学杀伤性武器的现实危害，以及它的任意性、不公平和对人的诉求的漠视。

但是，教师既是教育者和儿童看护者，也是员工，这里我想再从员工评估的角度探讨一下教师评估模型。我们来看一下蒂姆·克利福德的例子。他在纽约市的一所中学当英语教师，拥有26年的教学经验。几年前，克利福德得知自己被一个教师评估程序给出了低分评价，这个增值模型类似于导致萨拉·韦索基被开除的那个模型。在满分100中，克利福德只得了6分。

他感到很沮丧。“我不明白，我那么努力地工作，怎么可能会得这么低的分数。”他事后对我说。“说实话，在刚知道我的分数时，我感到羞耻，没有对任何人讲。但是，大概在一天后，我得知我们学校还有两名教师比我的得分还低，这给了我勇气告诉别人我的得分，因为我希望我的同事明白，他们不是特例。”

克利福德说，如果不是因为自己拥有终身教师资格，他在当年就会被学校开除。他说：“即使拥有终身教师资格，连续几年得到低分必定会对教师的职业生涯造成某种程度的负面影响。”此外，终身教师的低分记录会刺激教育改革者，视这种情况为就业保障部门正在保护不称职的教育者。克利福德因此为来年感到忧虑。

这一增值模型只是打出了一个低分，但没有给他任何有关改善的建议。所以，克利福德在下一学年采取了一如既往的教书方法，然后寄希望于奇迹发生。结果这一年，模型给他打了96分。

“你可能以为我会很开心，但是并没有。”他说。“我知道我以往的低分并不属实，所以也不会因为同一个有缺陷的程序给我高分而感到喜悦。得分反差这么大反而使我意识到，将增值模型运用到教育行业是多么可笑。”

“并不属实”这个词说得好。事实上，数据误读贯穿整个教师评估史。问题起源于《国家处在危险中》这份教育报告在分析教育现状时出现的大量统计错误。教育报告研究员基于一个本科生才会犯的根本性错误做出了一系列错误的判断。事实上，研究员自己对数据的错误解读就是美国教育失败的完美案例。

《国家处在危险中》发布7年之后，桑迪亚国家实验室的研究员再次分析了该份报告中的数据。这些研究员可都是建造和维护核武器的统计专家，他们很快就发现了错误。确实，高考平均分在下降。但是，在17年里，参加高考的总人数激增。大学的校门逐步对贫困学生和少数族裔学生开放，机会不断增多，这是社会良性发展的标志之一。这些新近获得考试机会的学生自然会因为准备不足拉低高考平均分。然而，当统计科学家把所有的高考学生按家庭收入进行分组后，他们发现，每一小组的高考分数都在提高，不管是穷学生组还是富学生组。

这种现象在统计学上叫辛普森悖论，即在某个条件下，两组数据在分别讨论时都满足了某种性质，可是一旦合并考虑，就可能导致相反的结论。教育报告《国家处在危险中》推导出的错误结论，催生了整个教师评估行动，而这一错误的结论来源于对数据的严重误读。

蒂姆·克利福德老师得到的反差极大的分数也是一个数据误读的结果，可见，数据误读经常发生。教师评分的根据是什么都体现不了的试卷。这听起来也许有点儿夸张。毕竟是学生在考试，而学生的考试成绩会影响克利福德的得分。这基本属实。但是，克利福德的两次得分，一次是丢人的6分，另一次是夸张的96分，都完全是基于靠不住的、近乎随意的近似值得出的。

问题在于管理层在追求公平的过程中忽略了精确度。他们明白，在富人学校中，医生和律师的子女上了精英大学，将这完全归功于教

师的教学是不对的；而对贫民区的教师评分也不该采用同样的考核标准，我们不能期待他们创造教学奇迹。

因此，管理层决定不用单一的标准来衡量教师，而是让模型去适应社会的不公：不去比较蒂姆·克利福德的学生与其他街区学校的学生，而是把他们的成绩和预测模型的预测成绩进行比较。每一个学生都有一个对应的预测分数。如果学生的成绩超过预测分数，教师就会得到表扬；如果低于预测分，教师就会遭到谴责。可能你会觉得这很荒谬，但相信我，我说的是真的。

从统计学上来说，在努力使测试不受阶级偏见和种族偏见影响的过程中，管理层所使用的模型从主要模型转变为次要模型。他们没有把教师评分建立在学生的直接测试上，而是根据所谓的误差项（实际结果和期望值之间的误差）给教师评分。在数学上，这是一个非常粗略的命题。因为期望值本身就是来源于数据的一个估计值，这相当于在猜测的基础上再猜测，其结果就是诞生一个会产生很多随机结果的模型，统计学家称之为“噪声”（noise）。

现在，你也许会认为大量的数据有利于模型做出更准确的预测，毕竟，纽约市的公立中学有110万名的学生，这一巨大的数据集足以让模型做出一些有意义的预测。如果有8万个八年级学生参加考试，那么，为贫民学校、中产阶级学校以及富人学校确定一个可靠的平均值是否可行?

可行。如果蒂姆·克利福德教过的学生总数很大，比如说1万，那么衡量这一届学生和上一届学生的平均分并从中推导结论就是合理的。数据的规模抵消了异常值的影响。从理论上来说，统计趋势在大样本中会表现得更明显。但是，拿一个只有25或30个学生的班级和更大的学生群体做比较几乎是不可能的。所以，如果一个班恰好有几类学习能力强的学生，那么这个班级的分数很快就会超过平均分。其他没有此类学生的班级的成绩则会进步得很慢。克利福德几乎完全不了

解这个不透明的数学杀伤性武器，但他认为，在此系统下得到的截然不同的两次评分就是源于学生类型结构不同的影响。在得分低的那一年，克利福德说："我教过很多特殊教育学生以及许多优秀学生。我认为我在那年评分极低，是因为班级中这两类学生所占比重比较大。特教学生的分数很难提高，因为他们有学习障碍，优秀学生的分数也难以提高，因为他们的分数已经很高了，进步空间很小。"

第二年，他的学生各式各样，但其中很多的学生都介于上述两类极端情况之间。克利福德的得分似乎表明他从一个失败的教师变成了一个优秀的教师。这样的结果非常普遍。教育工作者、博客博主加里·鲁宾斯坦研究发现，在连续几年教授同一门学科的教师当中，25%的教师在不同年份的得分有将近40分的反差。这表明评分数据实际上是随机的。不是教师的能力忽好忽坏，而是并不属实的数学杀伤性武器创造的评分系统导致了教师分数的巨大反差。

凭借着并没有实际意义的评估分数，增值模型造成了广泛而恶劣的影响。"我发现有些优秀教师根据得分说服自己最多只是个普通的教师，"克利福德说，"他们不再像以前那样教授很棒的课程，而是忙于应试。增值模型给出的低分会打击年轻教师的信心，同时让得分高的教师产生错误的成就感。"

正如其他许多例子中的数学杀伤性武器，教师评估模型的建立在最初也是出于好的意图。2001年，美国颁布《不让一个孩子掉队》教育法案，要求学生参加高风险的标准化考试。奥巴马政府早就了解到，该法案的受害者往往是贫穷学区的学校。所以，奥巴马政府不再要求能够自行证明教师教学质量的学区的学生参加上述考试，确保此类学校即使有学生掉队也不会因此受罚。注

教育评估模型的应用主要是源于针对这一法规所做的修订。但是在2015年年底，标准化测试热发生了大逆转。第一，美国国会和白宫

批准撤销《不让一个孩子掉队》法案，新的法案赋予各州更多的自主权，每个州将自主采取措施帮助相应学区扭转教学质量低下的状况。同时新法案也提供了更广泛的评价指标，包括师生互动，是否配备高阶课程，学校的学习气氛和校园安保情况。换句话说，教学管理官员要开始研究每一个学校的具体情况，而不再强调像教师评估增值模型这样的数学杀伤性武器给出的建议，或者不如说是最好完全取消后者。

在同一时期，纽约州时任州长安德鲁·库默的教育工作小组宣布根据学生考试成绩评估教师的机制将暂停4年。这一新规虽然很受欢迎，但是并不表示教育部门明确反对教师评估这个数学杀伤性武器，教育部门只不过是承认了它们确实不公平。这一新规的发布其实是得益于学生家长的推动，他们投诉标准化测试机制让他们的孩子疲惫不堪，一年中的上课时间太长。2015年春季，家长的联合抵制使得三年级到八年级的学生中有20%的人不用参加考试，这一数字还在增长。在和家长对话时，库默政府批判了教师评估增值模型。毕竟，如果大规模的学生不去参加那些标准化测试，纽约州就会缺乏普及这一评估机制的数据。

蒂姆·克利福德被这个消息所鼓舞，但仍然很担忧。“退出考试抗议活动威胁了库默的州长职位。”他在邮件中写道，“他担心失去优秀学区更富有选民的支持，这些选民过去一直是他最坚定的支持者。为了处理好这个问题，他才下令暂停了标准化测试机制。”

可能确实如此。而且，鉴于教师评估增值模式已经成为对抗教师工会的一种行之有效的工具，我认为它不会在短期内消失。这种评估模型的存在已经根深蒂固了，有40个州和哥伦比亚特区使用或开发了此类模型的其他形式。这为传播我在本书中对这些模型以及其他数学杀伤性武器的分析增加了理由。一旦人们识别了这些模型并了解这些模型的统计缺陷，他们就会要求对学生和老师都更公平的评估方法。

然而，如果一个模型的目标本来就是找个人承担责任，恐吓教师或者员工，那么，如我们所见，这一给出随机分数的数学杀伤性武器完美地实现了这个目标。

---

1. 《不让一个孩子掉队》教育法案规定，教学质量差的学校的学生可以转到教学质量好的学校。如果某校教学质量极差，政府会勒令该校关门，以特许学校取而代之。

第八章
替代变量和间接损害

信用数据的陷阱



小城镇的银行工作人员往往是大人物，因为他们掌控着金钱。如果你想买一辆新车或者按揭买房，你就得穿上最体面的衣服去拜访他们。而且由于银行家跟你生活在同一个社区，他很可能了解你的一些生活细节，比如你是否有做礼拜的习惯，你哥哥所有的违法经历，以及你的领导（也是他的高尔夫球友）对你的工作评价。理所当然，他也了解你的种族和民族身份，而且能看到你的贷款申请表上的得分。

不管是有意还是无意，上述这4个因素往往会影响这些“大人物”的判断。而且，他们通常更信任自己圈子里的人，因为人性就是如此。但是，对于数百万的美国人来说，这意味着前数字状态同我在本书所描述的某些数学杀伤性武器造成的不公平状态一样可怕。少数族裔和妇女等圈外人被习以为常地拒之门外。为了获取贷款，他们必须拥有一套完美的资历，然后还要找到一位开明的银行家。

显然，这不公平。而算法的出现改变了这种情况。数学家厄尔·艾萨克和他的工程师朋友比尔·费尔设计出了一个模型，他们称这一评估个人贷款违约风险的模型为FICO。该模型的唯一评分参数就是贷款者的资产，主要依据是贷款者的债务负担和账单支付记录。该模型的系统评分不受种族因素的影响。在数百万的贷款业务新客户不断涌入的情况下，该模型极大地促进了银行业的发展，因为该模型对风险的评估更加精确。FICO评分模型现在依然在被广泛使用。众多信贷机构，如益百利（Experian）、环联（Transunion）和艾奎法克斯（Equifax）等，都将各自的客户数据输入到FICO模型中以得出相应的分数。该模型有很多值得赞美的非数学杀伤性武器的特质。首先，该模型有清晰的反馈回路。信贷公司可以看到哪些借款人有贷款违约记录，因此信贷公司可以把借款人的违约数据和他们的得分进行比对。如果结果显示风险得分低的借款人的贷款违约行为比模型预测的更加频繁，则FICO和信贷公司可以对模型进行微调以增加其精确度。这是有益的数据利用方式。

其次，这种信用评分模型相对透明。比如，FICO的官网就向公众简单介绍了提高系统得分的方法（如减少债务，及时支付账单，不要频繁办理新的信用卡等）。同样重要的是，信用评分行业受政府管制。如果你对自己的评分结果有疑问，你有权向评分机构要求查看自己的信用报告，信用报告将包含所有的评分细节，比如你的按揭记录和效用报偿，你的全部债务以及目前的可用透支信用额度。虽然整个过程可能会慢得让人痛苦，但是如果你发现了什么错误，你可以修正它们。

自从比尔·费尔和艾萨克开创了此模型之后，信用评分系统的使用得到了广泛普及。如今，我们每一个人在统计科学家和数学家的眼中都是由各种各样的数据拼凑起来的总和，包括我们所在街区的邮政编码、上网模式以及最近的购物记录。许多伪科学模型也想要预测我们的信用可靠程度，试图给我们每个人一个所谓的电子评分。这些分数

在我们几乎不知情的情况下为我们中的一些人打开了机会之门，同时也把另一些人拒之门外。与FICO评分模型及其他类似的模型不同的是，电子评分更为武断任意，不负责任，不受管束，而且往往不公平。简单来说，电子评分是一种数学杀伤性武器。

位于弗吉尼亚州的中立星公司就是一个典型例证。中立星为客户公司提供客户锁定服务，包括为客户公司的客服中心提供客流量管理服务。该公司的一项技术可以用于快速处理来电者的所有可获得数据，并把来电者按一定的指标进行等级排序。排在前面的人被认为是更有价值的潜在客户，系统会迅速地把他们筛选出来，由人工客服接听电话。而那些排在后面的人要么需要等待很长的时间，要么被分配到机器人客服。

信用卡公司，如第一资本金融集团，对于浏览信用卡公司网页的访客，也有类似的快速算法，它们能够立刻获得该访客的网页浏览和购物模式的相关数据，由此深入了解其潜在客户。在汽车资讯网Carfax.com上点击新款捷豹汽车页面的人很可能比查询2003年款式的福特金牛座车型汽车的人更富有。大多数的评分系统还能够轻而易举地获得来访者电脑的IP地址。如果将该信息和房地产数据库进行匹配，则系统同样可以对来访者的财富水平进行推断。比起在冬奥克兰海湾上网的人，在对面的旧金山湾区Balboa Terrace社区上网的人显然是更有价值的潜在客户。

上述电子评分模型的存在不足为奇。我们已经讨论过仰赖类似数据的模型，如锁定穷人的掠夺式贷款模型和权衡黑人偷盗汽车概率的模型。不管怎样，这些模型都指引了我们进入一所学校（或者监狱），找到一份工作，以及在职场最大限度地发挥我们的价值。同样，当你准备买房子或买汽车的时候，资产评估模型自然要挖掘同样的数据对你进行评估。

但是，让我们来考虑一下这些电子评分系统产生的恶性循环。电子评分系统很可能会给来自东奥克兰偏远地区的借款者一个低评分。这个地区有很多人拖欠贷款。所以，从这些人的屏幕上跳出的信用卡办理广告将是那种锁定违约风险更高的人群的类型，这意味着向本已艰难谋生的人推荐可用透支额度更少、贷款利息更高的信用卡。

我们讨论的大部分掠夺式广告都是由这种电子评分系统生产的，包括发薪日贷款和营利性大学的广告。电子评分系统是信用评分系统的化身。但是，由于法律禁止企业使用信用评分进行市场营销，因此企业转而采用了这种不严谨的替代品。

这条对于信用评分的商用禁令有一定的逻辑。毕竟，我们的信用记录包含非常私人的信息，我们有权把控谁能看到这些数据，这很容易理解。但结果是，为了创建一个并行数据市场，企业开始深入挖掘一个庞大的、基本不受监管的数据池，比如点击流和地理标签。这个过程基本避过了政府的监督，企业得以根据效率增进、现金流和利润的增长来衡量自己的成功。除了极少数的例外情况，正义和透明这样的概念几乎不可能被纳入企业的模型算法之中。

我们姑且把这种算法比作20世纪50年代的银行家吧。这位银行家有意无意地权衡各种数据点，而这些数据点与他的潜在借款人是否能够承担抵押贷款可能没多大关系甚至毫无关系。换句话说，银行家只是扫视了一下自己的桌子，了解了潜在客户的种族，然后就据此得出了结论。客户父亲的犯罪记录可能是一个不利因素，而定期去教堂则可能被看作一个有利因素。

所有这些数据点都是替代变量。这位银行家本应在试图弄清楚一位客户的支付能力的过程中冷静地研究这些数据（一些传统的银行家一定会如此）。但是，与之相反，这位银行家草率地把这些数据与种族、宗教和家庭关系联系在了一起。此举避免了审查借款人的个人情况的麻烦，而把借款人归入一个群体之中，今天的统计学家称之

为“池”。然后，这位银行家会决定“像你这样的那类人”能不能被信任。

费尔和艾萨克最大的进步就是丢弃了作为“相关”金融数据的替代变量，包括关于过去的账单支付的行为模式。他们着重分析的是存在疑问的个人，而非有类似特质的一类人。相比之下，电子评分系统则让我们退回了过去。电子评分系统通过无数替代变量分析个人，在几毫秒内执行成千上万次“像你这样的那类人”的计算。如果结果显示“那类人”中有足够多的人是欠债不还者，或者更糟，是罪犯，那么最开始被评估的这个人就会得到相应的对待。

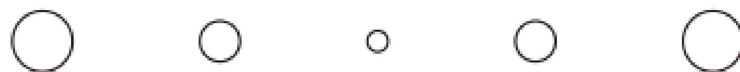
有时候，人们会问我如何教授数据科学家伦理观。我通常首先会让大家讨论如何构建一个电子评分系统模型，然后问他们把“种族”作为模型的输入数据是否合理。而他们必然会回答说这种做法不公平，而且很可能是非法的。我的第二个问题是问他们是否要在模型中使用“邮政编码”数据。乍一看，这一做法没什么问题。但是，这些人很快就会发现，他们把过去的不公正编码带进了新的模型中。当他们的模型将“邮政编码”数据考虑在内时，他们其实表达了这样一种观点：某个区域的居民的行为史可以决定，或者至少在某种程度上决定，住在那里的人应该得到什么样的贷款。

换句话说，电子评分系统建模者设法回答的是这个问题：“像你这样的那类人过去的行为表现如何？”而在理想的情况下，应该问的问题是：“你过去的行为表现如何？”

这两个问题的区别是巨大的。想象一下，如果一个积极进取、有责任心的移民白手起家，力图创业，而他需要仰赖这种系统获得初期投资，那么，谁会为这样的人冒险？基于人口统计学和历史行为数据建立的模型大概不会。

我应该提醒大家注意的是，在遍布替代变量的统计界，这种模型经常奏效。毕竟在很多时候，物以类聚，人以群分。有钱人买游轮和宝马，而穷人往往确实需要发薪日贷款。而且，由于这些统计模型在大多数情况下都奏效了，带来了效率提高，利润激增，因此投资者会加倍投资这些科学系统，让这些科学系统把成千上万的人归入正确的“池”中。这就是大数据的胜利。

那么，被误解、误放入错误的“池”里的人该怎么办呢？这种情况经常发生，而并没有可用的反馈回路用以修正系统。一个数据处理机器是不可能了解自己把一个有价值的潜在客户分配给了机器人客服的。更糟糕的是，被无监管的电子评分系统评选出的失败者无权抱怨，更不用说纠正系统的错误了。在数学杀伤性武器领域，他们的遭遇是附带损害。而由于整个黑暗的系统隐藏在遥远的服务器群中，他们几乎不可能弄清真相。他们中的大多数人可能会理所当然地得出这样的结论：生活就是不公平的。



在我刚刚描述的世界中，建立在数百万个替代变量上的电子评分系统在进行暗箱操作，而封装着个人信息和相关数据的信用报告则受到法律的保护。但遗憾的是，现实并没有那么简单。信用报告也经常作为替代变量出现。

社会上的众多机构，包括大公司和政府，都在寻找值得信赖的、可靠的人，这不足为奇。在讨论找工作的那一章里，我们看到公司会用软件筛选简历，将那些心理测试结果显示出具有不符合要求的个人特质的求职者拒之门外。另一种非常普遍的做法是考虑申请者的信用评分。雇主们会说，如果人们按时支付账单并避免负债，这难道不正意味着这个人可信又可靠？他们知道，这并不真的完全是一回事，但是他们认为两者的关联显而易见。

这就是使信用报告的应用范围扩张至远远超出其本来范围的例子。信用好已经成为其他美德的一个非常简单的替代品。相反，信用差则代表着许多与支付账单无关的罪恶和缺陷。正如我们所见，各类公司都将信用报告转化为本公司的评估模型中的信用评分，将其作为一个替代变量。这种做法既有害又无处不在。

对于某些特定的应用程序，这样的替代变量可能是无害的。例如，某些提供网上约会服务的网站会根据信用评分将用户配对。其中一个公司，“信用评分约会”公司（CreditScoreDating），在其宣传语中声明“好信用才是真性感”。我们可以质疑将财富与爱情联系起来的观点的合理性，但不管怎样，该公司的客户知道自己在做什么以及这么做的原因。是否购买这项服务取决于客户。

但是，如果你是在找工作，那么一次信用卡过期未还的记录或者曾经支付过学生贷款滞纳金的记录就极有可能对你不利。根据美国人力资源管理协会的一项调查，美国近一半的雇主会用求职者的信用报告筛选潜在雇员。也有一些雇主会查看在职雇员的信用状况，尤其是在职员提出晋升申请的时候。

公司在查看雇员的信用记录之前必须征得对方的同意，但这通常只是走个流程。在许多公司，那些拒绝交出信用记录的人根本不会被考虑录用。如果没有好的信用记录，他们就很可能被淘汰。2012年一份关于中低收入家庭信用卡债务的调查报告清楚地阐明了这一点。有10%的受访者表示，他们收到雇主的回复称，是他们的不良信用记录使他们丧失了这次工作机会。而谁都说不准究竟有多少求职者是因为自己的信用报告而被取消资格，却仍被蒙在鼓里的。尽管法律规定，当求职者因为信用不良被取消资格时，雇主必须明确向求职者说明，但大多数雇主只是简单地告诉这些因信用报告被淘汰的应聘者，他们不太合适这个工作，或者其他人更有资格。

利用信用评分指导招聘和升职的行为惯例导致了贫困的恶性循环。毕竟，如果你因为信用记录找不到工作，那么你的信用记录很可能会变得更糟，你找到工作的机会就会变得更小。这个问题和年轻人在第一次找工作时所面临的问题一样：因为缺乏经验而不被录用；或者和长期失业者面临的问题一样：他们发现几乎没有雇主会雇用他们，因为他们长期处于失业状态。对于那些本已陷入不幸境地的人来说，这种情况就是一个不断升级的恶性循环。

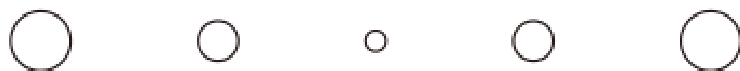
当然，雇主对这种说法不以为然。他们认为，好信用是一个靠谱的人的基本特质，他们想要聘用的正是这样的人。但是把欠债看作道德问题就是一个错误。每天都会有很多努力工作且值得信赖的人因为公司破产、削减成本或者向境外转移工作等原因失去工作。而且在经济萧条时期，这部分失业者的人数还要更多。很多人刚失业就没有了医疗保险。在这个节骨眼上，万一发生点儿事故或者生一场病就很可能会导致他们错过还款期限。即使奥巴马推出了《平价医疗法案》，扩大了医保范围，但医疗费用仍然是破产的一个最大根源。

有储蓄的人当然可以在经济萧条时期维持信用记录的完好，而月光族则更易遭受冲击。因此，信用好不仅仅被视为负责任和明智的替代变量，也被视为财富的替代变量。而财富和种族高度相关。

思考一下下面这种情况。2015年，美国白人家庭的平均财产和资产大约是黑人家庭和拉美裔家庭的10倍。仅有15%的白人的资本净值为零或者资本净值为负，而超过1/3的黑人和拉美裔家庭没有储蓄。这种贫富差异随着时间的推移还在扩大。到这些人60岁的时候，白人将比非裔及拉美裔美国人富有11倍。鉴于以上这些数字，不难证明由雇主查看信用记录导致的贫困陷阱会阻碍社会公平和种族平等的实现。本书写到这里的时候，美国已经有10个州颁布法律判定利用信用评分指导招聘属违法行为。在颁布该项法律时，纽约市政府指出，查看信

用记录“不成比例地影响了低收入求职者和非白人求职者”。目前，利用信用评分指导招聘的做法在美国其余的40个州仍是合法行为。

我并不是说全美国的人事部刻意地制造了一个贫困陷阱，甚至一个种族陷阱。他们毫无疑问是真的相信信用报告里面有支持他们做出重要决策的相关数据。毕竟，“数据越多越好”是信息时代的指导原则。但考虑到社会公平，一部分数据理应被排除在外。



想象一下这个场景：你从斯坦福大学法学院刚毕业，要面试旧金山的一家知名律所。律所的高级合伙人看着一份电脑生成的个人分析报告大笑起来——这份报告显示你曾因为运营一个制毒实验室被逮捕过！他觉得不可思议，然后果断淘汰了你。而事实是，你的名字是个常用名，这份报告的出现肯定是因为电脑犯了个愚蠢的错误，把你的名字同别人的搞混了。但面试就这样继续下去了。

富人往往能仰赖更具个性化的软件做出重大决定；但是对普通阶层的人而言，尤其是较低阶层的人，他们的工作中的大部分操作都是纯自动化的。如果历史档案有错误（这是常有的事），即便再精良的算法也不可能给出正确的决策建议。数据大牛早就说过：无用输入，无用输出。

位于自动操作程序接收端的人会因为一个错误的决策受苦多年。比如说，大家都知道，由电脑生成的恐怖主义者禁飞名单就常常制造错误。一个无辜的人，可能只是因为名字和某个恐怖主义者相似，在每次登机的时候就可能面临地狱般的折磨。（与之相反，富有的旅客往往能够花钱购买“可靠乘客”的身份，因此得以顺利通过安检。实际上，他们花钱购买的就是一个可以避免数学杀伤性武器伤害的防护盾。）

类似的错误随处可见。2013年联邦贸易委员会发布的一篇报告显示，全美约有5%的消费者（约1000万人）其信用报告中包含差错，而差错带来的后果足以严重到增加他们的借贷成本。这个问题很麻烦，不过至少，信用报告存在于数据经济规范化的一面。消费者可以并且应该每年要求查看自己的信用报告，修正可能发生的昂贵错误。[注]

而数据经济不规范的一面破坏性更强。许多公司，不论是像安客诚这样的大型数据管理公司还是众多的小数据贩子，都会从零售商、广告商、智能手机应用开发商以及博彩公司或者社交网站公司购买用户信息，收集每一个消费者的大量相关数据。比如说，他们可能会标明某个消费者是否有糖尿病，居住环境里是否有吸烟者，是否开越野车，以及是否有一对牧羊犬（这对牧羊犬可能在死后很长时间还会保留在主人的档案里）。这些公司还会收集各种公开的政府数据，包括选举信息、逮捕记录以及住房销售记录等。所有这些数据组成了一个消费者的档案，又被这些数据公司出售给他们的客户。

当然，部分数据代理商还是比较可靠的。但是，任何企图根据数千个不同的信息来源对数百万的人进行档案概括的操作，都可能出错。以费城人海伦·斯托克斯的遭遇为例。她想搬到一个当地的高级生活社区，但她因为档案里的逮捕记录屡次遭到拒绝。的确，因为与前夫发生争执，她曾两次被逮捕。虽然她并没有被定罪，而且这些记录应该已经从政府数据库中被删除了，但由RealPage数据管理公司整理出的她的档案里仍有这项逮捕记录，这家公司的主要业务是向房地产公司提供对房客的背景调查。

建立并出售个人档案为RealPage等同类公司创造了收入。海伦·斯托克斯这样的人并不是它们的客户，而是它们的产品，应付这些人的投诉会浪费精力和财力。毕竟，虽然海伦·斯托克斯说她的逮捕记录已经被删除了，但核查这个事实也会耗费时间和金钱。有些人可能只好通过在互联网上发送几条信息甚至打个电话（希望不必如此）来解决

这个问题。毫无悬念，斯托克斯的逮捕记录始终没有被删除，直到她提起上诉。而且，即使RealPage公司解决了这个问题，谁知道还有多少其他的数据代理商会继续贩卖包含同样错误信息的档案呢?

的确，有些数据代理商会给消费者提供数据查看权限。但是这些数据报告是被组织过的。报告里确实包含部分事实，但并不总是包含数据经纪人根据算法从数据中得出的结论。比如，若某人想方设法拿到了查看自己在某个数据公司的档案的权限，她可能会发现档案中包含她的住房抵押、威瑞森通信公司的账单以及459美元的车库修理费等记录。但是她不会看到自己身处一个被命名为“乡下人，勉强维持收支平衡”或者“老来无退休收入”的群体分类中。对数据代理商来说，幸运的是，很少有人有机会看到这些细节。如果我们看到了这些细节，而美国联邦贸易委员因此向数据代理商追责，那么数据代理商很可能会被来自数百万消费者的投诉围攻，这将严重破坏它们的商业模式。但就目前看来，消费者常常只是在无意之中了解到自己的档案有差错。

举个例子：美国阿肯色州的一个居民，名叫凯瑟琳·泰勒，她在几年前错失了一个当地红十字会的工作机会。这件事本身没什么稀奇的。但是，泰勒收到的拒绝信上含有一条非常“有价值”的信息：她的档案报告里有一项她因有意图制造并且销售毒品而遭到刑事控告的记录。显然，这样的人不会是红十字会想要聘用的人。

泰勒研究了一番，发现遭到刑事控告的是另一个与她同名同姓的人，而且这个人刚好跟她在同一天出生。她后来发现，至少还有十家公司的档案报告包含这条错误的、严重损害她的声誉的信息，其中一份档案还导致她在申请联邦住房补贴时被拒，她本来没有意识到这才是她的申请被拒的真正原因。

在完全自动化的程序中，这种情况无疑很有可能会发生并造成无数恶果。但是在泰勒的例子中是有人力参与的。在重新申请联邦住房补贴的时候，泰勒和她的丈夫会见了房管局的一位职员以完成他们的

背景调查。该职员名叫旺达·泰勒（和凯瑟琳·泰勒无亲属关系），她使用的信息是由数据代理商Tenant Tracker提供的。这份数据错误百出，身份信息混乱。报告不仅把泰勒和可能是某一个刚好与她出生在同一天的、可能是以钱特尔·泰勒作为化名的重罪犯搞混了，还把她和另一位她已经听说过的那个凯瑟琳·泰勒搞混了，后者在伊利诺伊州因为偷盗、伪造以及持有管制药物而被定罪。

总而言之，整个档案报告极其混乱。但是，房管局的旺达·泰勒经历过同样的事情，所以她开始着手研究这件事。首先，她立即把报告中被推定为可能是泰勒化名的钱特尔的相关信息划掉了。之后，她在档案报告里发现，那个伊利诺伊州的小偷的脚踝上纹有一个名字，“Troy”。在检查完凯瑟琳·泰勒的脚踝之后，她把这个罪犯的相关信息也删掉了。这次会面结束之后，这位细心的房管局职员把由数据收集程序带来的错误理清楚了。现在，房管局已经清楚地知道这个住房补贴申请者是哪一个凯瑟琳·泰勒了。

留给我们的问题是：会有多少个细心的旺达·泰勒去纠正我们数据中的错误呢？答案是：并不多。数据经济环境下，大部分人要么是局外人要么是老古董。各种系统的开发都以尽可能使其自动化运转为目标。这是一种高效率的方式，也是利润的来源。和所有其他的统计程序一样，错误是不可避免的，但是，减少错误的最快捷方式是微调机器运转算法。人类只会把事情搞砸。

由于计算机能够理解越来越多的书面语，还能按照某些要求快速处理成千上万的书面文件，如今，自动化呈现飞速发展的趋势。但是，计算机仍然会犯各种各样的错误。IBM超级计算机系统“沃森”挑战了美国智力问答游戏“Jeopardy!”，尽管总体表现优秀，但还是有10%的时间，其不能理解问题的描述语言或者语境。“沃森”说蝴蝶的饮食是“犹太食品”，还把查尔斯·狄更斯的小说《雾都孤儿》中的主人

公奥利弗·特维斯特和20世纪80年代的流行电子乐乐队“宠物店男孩”搞混了。

这些错误肯定大量存在于我们的档案之中，而充斥着混淆和误导的算法正日益掌控着我们的生活。这些由自动化数据收集程序带来的错误正在污染预测模型，助推数学杀伤性武器的诞生。并且此类充满错误的数据收集程序只增不减。计算机功能已经拓展到了理解书面语，在理解口语和图片方面也取得了很大的进展，借助技术的进步，计算机正致力于捕捉宇宙中一切事物的相关信息，包括我们人类。这种新科技将为扩充我们的档案挖掘更多的新信息，这同时也增加了犯错的风险。

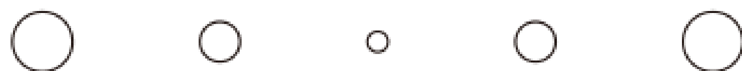
2015年，谷歌的图片自动识别软件将照片中三个年轻的美国黑人标记为黑猩猩，谷歌为此郑重道歉。但是在类似的系统中，此类错误的发生是不可避免的。很可能是错误的机器学习（而不是谷歌总部某个种族主义者的任意行事）导致计算机把现代智人和我们的近亲黑猩猩相混淆的。识别软件学习了数十亿张灵长类动物的图片，然后自行发展出了一套区分的方法：以皮肤色度、两只眼睛之间的距离以及耳朵的颜色作为主要判断指标。但是显然，谷歌的图片识别软件在发布前没有进行更为全面的检测。

这些错误为机器创造了进一步学习的机会，前提条件是系统能接收到错误反馈。谷歌的识别软件接收了错误反馈。但是不公平仍然存在。在自动化系统仔细查看我们的数据，对我们进行评估并给出一个电子评分的过程中，系统自然会把我们的过去投射到未来。就像我们在再犯模型和掠夺式广告模型中已经看到的，穷人会被认为一直贫穷，并因此得到相应的对待：被机会之门拒之门外，更多地被逮捕，不能享受平等的服务和贷款条件。这是一种冷血的做法，不透明且受害者无法申诉，没有任何公平可言。

但是，我们不能指望自动化系统自行解决这个问题。尽管机器功能强大，但它们目前还不能照顾到公平，至少它们自己不能。筛选数据并且判断什么是公平对它们来说是完全陌生且极其复杂的。只有人类能够给系统施加与公平相关的限制条件。

矛盾出现了。如果我们最后一次回顾一下20世纪50年代的银行家，我们会发现他的大脑充斥着各种人类的劣根性，包括欲望、偏见和对外来者的不信任。而正是为了更加公平有效地开展工作，他和行业里的其他人把工作交给了算法。

60年以后的今天，世界被自动化系统控制，这些系统所做出的判断仰赖的是我们漏洞百出的数据档案。系统迫切需要理解语境、常识以及只有人类才可以提供的公平。但是，如果我们把这个问题交给关注效率、增长以及现金流（并且可以容忍某种程度的错误）的市场，想要干预的人类将会被命令远离系统。



这是个难题，因为即使我们的旧信用模型变得透明了，强大的新系统仍然在不断涌入市场，试图取而代之。比如，脸书就发明了一款建基于人们在社交网络上的行为数据的信用评级软件。该软件的表面目标是合理的。想象一个花了五年的时间投身于把饮用水引进非洲贫困地区这项公益事业的大学毕业生。回到家后，他没有信用评级，贷款困难。但是他脸书上的同学有投资银行家，专业领域的博士，还有软件设计师。“物以类聚，人以群分”的理论表明他是可以信任的。但是，同样的理论很可能会对伊利诺伊州东圣路易斯地区勤劳的房屋清洁工不利，因为这位清洁工很可能有很多失业的朋友，甚至还有几个正在坐牢的朋友。

与此同时，银行业正为了促进业务增长疯狂搜刮个人数据。但是，持牌银行受联邦法规和信息披露制度的管束，这意味着通过挖掘数据建立客户档案具有信誉和法律上的风险。美国运通公司在2009年经济大萧条时期就付出了惨痛的代价才吸取了这一教训。想要减少公司资产负债表上的风险项，美国运通无疑需要降低某些客户的消费上限。但是，与电子评分经济环境下的非正式金融机构不同的是，信用卡巨头公司必须要向客户解释原因。

美国运通公司想出了一个损招。公司宣称，在某些机构购物的持卡人更有可能拖欠还款。这是个统计结论简单直白，说明的是购物模式和信用违约率之间存在明显的关联。美国运通公司让它的客户自行猜测到底是哪一个机构破坏了他们的信用：到底是每周在沃尔玛超市的购物，还是在机械修理厂进行的刹车修理，使他们被归类为潜在的贷款违约高风险群体?

不管出于什么原因，这些人的信用因此被降低。更糟糕的是，消费上限被降级的记录几天之内就会出现在他们的信用报告上。实际上，可能早在他们收到通知之前，这条记录就已经出现在他们的信用报告上了。这将导致他们的信用评分降低，抬高他们的贷款成本。可以确定地说，这些持卡人中有很多人会经常去逛“信誉不良的穷人开设的商店”，因为他们没有很多钱。总之，算法注意到了这件事，而这让他们因此变得更加贫穷。

持卡人的愤怒得到了包括《纽约时报》在内的主流媒体的关注，于是，美国运通公司立即宣布不再将特定购物商店和信用风险相关联。（美国运通公司后来一口咬定是在通知中选错了用词，其声称公司只审查了广泛的消费模式，并没有针对具体的商店做进一步的分析。）

这件丑闻令美国运通公司头疼不已。即便公司确实发现了购物商店和信用风险之间存在密切的关联，公司现在也没法利用这项发现进

行筛选了。比起互联网经济下的大多数公司，它们更受法规约束。（但我并不是说运通公司应该对此提出抱怨。几十年来，公司历任总裁的说客说服政府起草了很多法规以维护公司根深蒂固的权威，试图将讨厌的行业新来者关在门外。）

所以，金融行业的新来者会选择更自由、更不受管束的路径并不奇怪。毕竟，创新仰赖自由实验。因此，在拥有海量的行为数据，且几乎不受政府监管的前提下，这些新来者创造新的商业模式机会多多。

比如说，很多金融公司致力于抢占银行的发薪日贷款市场。这些银行迎合了贫穷的工人阶级的需求，洗劫了他们每个月的工资，并索取高额利息。借款22周以后，500美元的贷款可能变成1500美元的负债。所以，如果有高效率的新公司能够找到评估还款风险的新方法，救信用可靠的贷款申请者于发薪日贷款的水深火热之中，向他们提供利率比发薪日贷款稍低一些的贷款，那么它也能挣到很多钱。

这就是道格拉斯·梅里尔的主意。梅里尔是谷歌公司的前首席运营官，他相信他能使用大数据技术计算风险，以相对低的利率提供发薪日贷款。2009年，他创立了一家金融公司，ZestFinance。在公司官网上，梅里尔宣称“所有数据都是信用数据”。换句话说就是，任何数据都会在贷款申请中起作用。

ZestFinance公司购买那些能够表明申请者是否曾经拖欠话费的数据，还有其他大量可获得的公开数据或者购物数据。梅里尔承诺，公司所发放贷款的利率将低于发薪日贷款商索取的利率。从ZestFinance公司贷款500美元，22周后，贷款人需还款900美元，比发薪日贷款的还款标准低了60%。

这确实是一种改进，但是，这样公平吗？ZestFinance公司的算法要处理每个申请者的高达一万个数据点，包括一些不寻常的观测值，

比如申请者的贷款申请表是否正确使用了大小写，他们需要多长时间阅读这份申请表，以及申请者是否仔细阅读了合同条款。ZestFinance公司认为，“规则遵守者”的信用风险更低。

这也许是对的。但是标点符号和拼写错误也表明低教育水平，而低教育水平和阶级、种族高度相关。所以，当穷人和移民有资格申请贷款时，他们不合标准的语言水平会抬高他们的贷款利率。如果他们后来因为高利率而还款困难，这可能又会反过来证实最开始将他们归为高风险群体的假设，他们的信用评分也将因此降低。这是一个恶性循环，而按时支付账单起到的作用不大。

这些公司的初心可能是好的，但当它们基于数学杀伤性武器发展壮大起来之后，麻烦也就随之而来。以P2P借贷公司为例。P2P最近十年刚刚兴起，其为借款人和贷款人提供了一个进行直接交流的中介平台。这本应表明的是银行业的民主化——更多的人可以得到贷款，与此同时，数百万的平民将成为小型银行家，得到更多的额外收入。借贷双方都能避开贪婪的大银行。

2006年，P2P鼻祖“借款俱乐部”公司（Lending Club）成立了。作为脸书网站上的一个小应用，其在一年后获得融资，变身成为一种新型银行。为了计算借款人的风险，借款俱乐部把传统的信用报告和全网搜集来的数据混合在一起。总之，其开发的算法也生成了一个电子评分，公司声称其给出的电子评分比信用评分更加精确。

借款俱乐部和其最大的竞争对手Prosper目前仍属于小型公司。这两家公司的贷款额度还不到100亿美元，在价值3万亿美元的借贷市场中只是沧海一粟。但是，这两家公司吸引了密切的关注。花旗银行和摩根士丹利的总裁加入了多家P2P公司的董事会，富国投资基金则成为借款俱乐部最大的投资人。2014年12月，借款俱乐部公开上市，打造了当年资金规模最大的科技公司IPO，公司总融资金额达到8.7亿美元，市值达90亿美元，在美国银行的市值排名中高居第15名。

但这场金融界的狂欢与资本的民主化和撤销中间人没有多大关系。据《福布斯》杂志的一篇报道，大型机构投资资金现已占到P2P平台所有资金的80%以上。对于大银行来说，这种新的借贷平台是监管严苛的银行业务的便捷替代品。在P2P系统中，借款人几乎能够分析其选择的每一个数据并据此给出他自己的电子评分。P2P平台完全允许借款人建立社区、街区和购物商店的风险相关性，而且借款人在做这些事情的时候完全不用给贷款申请者发送尴尬的信件解释原因。

那么，这对于我们来说意味着什么？随着电子评分甚嚣尘上，我们被一些秘密算法归类分组，其中有些算法仰赖的还是错误百出的个人档案。我们不是被当作个体，而是被当作某个群体的一员，被迫戴上了某顶帽子。电子评分污染了金融行业的大环境，贫民的机会越来越少。实际上，比起众多胡作非为的数学杀伤性武器，过去那种怀有偏见的银行家看起来也没有那么坏了。至少，过去的借款人能面对面地与银行家交流，并将自己的诉求向对方清晰地表达出来。

---

1. 需要指出的是，即便如此，修正这一错误仍然可能变成一个噩梦。一位密西西比州的名叫帕特丽夏·阿默尔的居民，用了两年的时间才让益百利信贷公司从她的信用报告上把她从来没借贷过的40000美元债务撤销。她的努力一直没有起效，直到她给密西西比的首席检察官亨利打了个电话。

第九章
“一般人”公式

沉溺与歧视



19世纪晚期，德国著名统计学家弗雷德里克·霍夫曼制造了一个潜在的数学杀伤性武器。霍夫曼当时就职于普天寿保险公司，我相信他当时建立这个模型并不是出于恶意。

后来，他的研究为公众健康做出了巨大贡献。他在疟疾防治方面做了很多有价值的工作，并且是最早发现癌症与烟草有关的人之一。然而，在1896年的春天，霍夫曼发表了一份长达330页的报告，这份报告严重推迟了美国种族平等的实现，固化了数百万黑人作为二等公民的地位。他的报告使用详尽的统计数据说明了美国黑人的生活很不稳定，并指出从保险业的角度讲，这一整个种族都无法投保。

和我们讨论过的许多数学杀伤性武器一样，霍夫曼的分析在统计学上是有缺陷的。他混淆了因果关系和相关性，所以他收集的大量数

据只证明了他的一个论点：种族是预期寿命长短的决定性因素。但种族主义已经深深根植于他的大脑，他显然从未停下来思考这些问题：贫穷和不公与美国黑人的高死亡率是否有关？缺少好的学校、现代化的卫生设施、安全的工作场所和医疗保险是否会导致黑人在更年轻的年龄死亡？

霍夫曼还犯了一个根本的统计错误。就像1983年美国教育局发表的教育报告一样，霍夫曼忽略了对结果进行细分。他认为，黑人只是一个庞大的同类群体，而没有把黑人按照不同的地理、社会或经济群体进行分类。他认为，一位在波士顿或纽约有稳定生活的黑人教师和一个在密西西比三角洲每天赤脚工作12小时的黑人佃农没什么区别。

他所在的保险行业也持有同样的观点。当然，随着时间的推移，保险公司的观念也会进步，它们开始向美国黑人家庭出售保险。毕竟，这里有钱可赚。但它们在几十年里一直相信霍夫曼的想法，即整个黑人群体的风险都更大，其中一部分则属风险极大行列。保险公司和银行会圈定它们不愿意投资的社区。这种冷血的做法现在被称作“地域经济歧视”（redlining），目前已经有多项法律条款予以明确禁止，包括1968年颁布的《公平住房法》。

然而，近半个世纪后，地域经济歧视仍然存在，只不过以更为微妙的形式被编码进了最新一代的数学杀伤性武器中。这类新模型的设计人和霍夫曼一样混淆了相关性和因果关系。新一代数学杀伤性武器压榨穷人，特别是少数族裔，它们用大量的统计数据支持自己的分析结论，还为自己赢得了公平科学的好名声。

在这趟被算法支配的人生之旅中，我们好不容易进入了一所大学，找到了一份工作（即使是一份需要我们随叫随到的工作）。我们得到了贷款，也看到了我们的信用评级是如何代表我们的其他美德或恶习的。现在，我们该去保护我们最宝贵的资产，即我们的家、车以及家人的健康，并为那些我们总有一天会离他们而去的人做安排。

保险是由精算科学发展而来的，这门学科的起源可以追溯到17世纪。当时，欧洲的资产阶级掌握了巨额财富，很多富人第一次开始为后代考虑。

数学的进步为预测提供了必要的工具，借助这些工具，早期的数据挖掘者开始寻找新的可供统计的数据。其中一位数据挖掘者是一位来自伦敦的服装商，名叫约翰·格朗特。他研究了出生和死亡记录，并于1682年发表了有史以来第一篇关于一整个社区的死亡率的研究报告。例如，他计算出伦敦的儿童在出生后的头6年里，每一年都面临着6%的死亡风险。（据此统计数据，他得以破除了一个当时的神话，即每当有新君主即位，瘟疫就会蔓延。）数学家们第一次计算出了一个人一生中最具可信度的一条预测曲线。当然，这些数字并不适用于个人，但如果有足够多的数据，平均值和浮动范围就是可以预测的。

数学家们并没有假装自己可以预见每一个个体的命运。个体命运是不可知的。但是，他们可以预测在一大群人中发生事故、火灾和死亡事件的频率。在接下来的三个世纪里，潜力巨大的保险行业借助这些预测兴起了。保险这一新行业第一次给了人们一个机会拿他们所在群体的集体风险为赌注，换取不幸发生时对自己的保护。

现在，随着数据科学和联网计算机的发展，保险业面临着根本性的变化。随着可获得的个人数据越来越多，包括我们的基因组，我们的睡眠、锻炼和饮食的模式，以及我们开车的熟练程度等，保险公司逐渐获得了精确计算个人风险的能力，并将自己从对更大的群体的集体风险的赌注中解放出来。许多人认为这是一个可喜的变化。如今，一位健康爱好者可以用数据展示她每晚睡8个小时，每天步行10英里，只吃一些绿色蔬菜、坚果和鱼油。难道她不应该暂停她的健康保险吗?

我们将会看到，这个行业的个体化转变尚未成熟。但是，保险公司已经开始利用数据将我们分成更小的群体，以不同的价格给我们提

供不同的产品和服务。有些人称之为定制服务。问题是，这并不是真正针对个体的服务。模型在我们看不到的地方仍然把我们归类为各种各样的群体，以各种行为模式为指标。不管最终的分析正确与否，这种不透明性都会导致欺诈。

以汽车保险为例。2015年，《消费者报告》杂志的研究人员在全美范围内开展了一项广泛的调查，旨在寻找保险价格的差异。他们为被设定为来自全美33419个地区的虚拟消费者分析了所有主要保险公司共计20多亿美元的保险报价。研究结果显示，保险公司给出的报价非常不公平，而且正如我们在上一章所见，这种不公平的报价是基于信用评分做出的。

保险公司从信用报告中获得这些分数，然后利用保险公司自己开发的算法创建自己的评级系统或者电子评分，这些评分就是驾驶可靠性的替代变量。但《消费者报告》发现，电子评分，包括司机的各种人口统计数据，往往比司机的行车记录权重更高。换句话说，在汽车保险公司的眼中，你管理金钱的水平要比你的开车水平更重要。以纽约州为例，当一名司机的信用评级从“优秀”下降到“好”时，其汽车保险费就会上升255美元。在佛罗里达州，拥有良好的驾驶记录和低信用评分的成年人平均比那些信用评级为“极好”但有酒后驾车记录的司机平均每年多支付1552美元的汽车保险费。

我们已经讨论了信用评分的广泛应用为何对穷人的不利程度更大。汽车保险费用就是这一现象的又一个例子，尤其是因为汽车保险对任何司机都是强制性的。而这个例子的不同之处是，在有直接相关的数据可用时，模型仍然关注的是替代变量。对于汽车保险公司来说，我想象不出有比酒后驾驶记录更有意义的数据了。这正是它们想要预测的领域的风险性的直接证据，比它们所考虑的那些替代变量要有效得多，比如什么高中时期的平均绩点之类。然而，驾驶记录在保

险公司的算法中远远没有从信用报告（我们已经知道，信用报告有时候也会出错）中拼凑得出的金融数据权重高。

那么，为什么汽车保险公司的模型更关注信用评分呢？就像其他的数学杀伤性武器一样，自动化系统可以高效地、大规模地进行信用评估。但我认为，更主要的原因还是利润。如果一个保险公司有一个系统，它可以每年从一个有着清白记录的司机身上额外获取1552美元的收益，那为什么还要改变它呢？正如我们在本书其他章节所见，数学杀伤性武器的受害者更可能是穷人和受教育程度低的人，他们中可能有许多人是移民，他们不太可能知道自己正在被保险公司压榨。而且，在他们生活的社区里，发薪日贷款商比保险公司还多，因此，找到一个保险费更低、更合理的公司对他们而言就更难了。简而言之，虽然电子分数可能与安全驾驶没多大关系，但它确实创造了一个利润丰厚的弱势司机群体。他们中的许多人极度渴望开车，因为他们需要以此谋生。对他们收取高昂的保险费用有助于提高公司的利润。

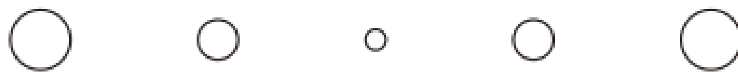
站在汽车保险公司的角度，这是一个双赢的结果。成功让一个信用差的好司机来投保是一项低风险高回报的投资。而且，保险公司还可以利用这部分收入解决由公司模型中的低效环节引起的问题，包括为那些信用好的坏司机支付酒驾撞车的保险赔偿金。

这听起来有点儿愤世嫉俗，但思考一下自诩为“人民的好帮手”的保险公司好事达的定价优化算法。根据美国消费者联合会的数据，好事达通过分析消费者的人口数据预测消费者购买低价商品的可能性。如果可能性较低，那么向他们收取更高的费用就是合理的。这就是好事达在做的事。

还有更糟糕的事情。在一份美国消费者联合会向威威斯康星州保险部提交的指控书中，其列举出了好事达公司的定价机制中的10万个微区段。这一定价等级系统是基于每一区段的消费者的预期支付价格建立的。因此，有些人会得到低于平均价格90%的巨大折扣，而有些

人则面临着800%的加价。“好事达公司保险的定价违反了基于风险收取保费的法规。”美国消费者联合会的保险部主任、前得克萨斯州保险局局长罗伯特·亨特表示。好事达公司回应称，美国消费者联合会的指控并不准确。不过，该公司也承认，其“曾认为在定价时考虑市场因素，这与行业惯例是相一致的，也是合适的”。换句话说，好事达的模型研究了大量的替代变量以计算可以向客户收取多少费用，而保险行业里的其他公司也是这么做的。

由此模型产生的定价必然是不公平的。如果保险定价是透明的，并且消费者可以货比三家，上述这种情况就不会发生。但和其他数学杀伤性武器一样，保险公司的模型也是不透明的。因为每个人都有不同的经历，而这些模型的设计目标是从有迫切需求以及对其操作方式一无所知的人身上赚取尽可能多的钱。结果是又一个恶性循环诞生了，那些尚能支付离谱保险费的可怜司机的每一分钱都被压榨了。对这个模型进行的微调也是为了进一步从这个群体中赚取尽可能多的钱。结果，他们当中的一些人在汽车贷款、信用卡还款或租金缴纳等方面不可避免地违约了。他们的信用评分因此进一步降低，这毫无疑问将使他们被划分至一个更绝望的细分群体中。



《消费者报告》发布了一份揭露汽车保险业黑暗面的报告，同时启动了一个针对美国保险监督官协会（NAIC）的活动，包括推特活动：@NAIC_News to Insurance Commissioners: Price me by how I drive, not by who you think I am!（致保监会保险官：收取费用请参考我的开车技术，而不是参考你想象中的我。）#FixCarInsurance（改善汽车保险）。

这项活动的宗旨是保险公司应该根据司机的驾驶记录，即他们的超速罚单数量，以及他们是否出过交通事故等判断他们的驾驶风险，

而不是根据他们的消费模式、他们的朋友或邻居来判断。然而，在大数据时代，说服保险公司根据我们的驾驶技术判断我们的驾驶风险性还没有过先例。

保险公司现在有各种各样的方法可以用来仔细研究司机的行为。只需要看看卡车运输行业就知道了。

如今，许多卡车都携带着一种电子记录装置，它能记录卡车的每一次转弯、每一次加速，每一次刹车。2015年，美国最大的货运公司转运交通公司开始在车内安装两个方向的摄像头，一个指向前方的道路，另一个朝向司机。

这种监视的目的是减少事故。每年大约有700名卡车司机死在美国公路上。这些车祸同时也夺走了许多其他车辆的司机的生命。除了个人悲剧之外，卡车造成的事故也带来了巨大损失。根据美国联邦汽车安全管理局的数据，一场致命车祸造成的经济损失约为350万美元。

但是，对于拥有庞大的分析实验室的卡车运输公司来说，保障运输安全只是它们的任务之一。如果你把地理位置信息、车载跟踪技术和摄像头结合起来，卡车司机将源源不断地提供大量的行为数据。现在，卡车运输公司可以分析不同的路线，进行燃料管理，以及比较白天和晚间不同时间段的运输效率。它们甚至可以计算不同路面的理想行进速度，并利用这些数据算出哪些运输模式能实现以最低的成本换取最高的收益。

它们也可以比较司机个体的表现。分析仪表板给了每个司机一张记分卡。只需点击一两次，管理者就可以在数据库中找到表现最好的和表现最差的司机。当然，这些监控数据也可以被用于计算每个司机的驾驶风险。

这个承诺并没有在保险行业中消失。前进保险、州立农业保险和旅行者等大保险公司都向同意分享驾驶数据的司机提供了保险费折扣。汽车里的一个小型遥测装置，就是飞机上的黑匣子的简化版，它将记录汽车的速度，司机如何刹车和加速等各项数据。另外，还有一个GPS监控将被用于跟踪汽车的行进线路。

从理论上讲，这符合《消费者报告》发起的活动的理想目标：让司机个体成为关注焦点。我们考虑一下刚到18岁的司机群体。传统上，他们需要支付极高的保险费，因为该年龄段的司机群体在统计上有更多的鲁莽行为。但现在，如果一名高中生从来不进行快速启动，能在限速范围内保持匀速前进，且会减速等红灯，那么他的保险费就会被打折。保险公司长期以来一直都偏好已经通过驾照考试的年轻司机，因为它们认为这是靠谱驾驶的替代变量。但是驾驶数据更真实、更直接，它应该比替代变量更好，对吧?

这里有几个问题。首先，如果系统在评估风险时将地理位置因素纳入考量，那么那些贫穷的司机就被优惠的保险项目排除在外了，因为他们更有可能在保险公司认为危险的地区开车，而且其中的许多人都是通勤时间长、上班时间不规律的上班族，这又进一步增加了他们的驾驶风险。

你可能会说这没什么错。如果贫困地区的驾驶风险更大，尤其是汽车盗窃的风险大，那保险公司为什么要忽视这些信息呢？如果通勤距离长会增加事故发生的可能性，这也是保险公司有权考虑的问题。这个判断仍然是基于司机的行为做出的，而不是根据他的信用等级或他这个年龄段的驾驶记录概况等无关替代变量做出的。许多人会认为这是一种进步。

在某种程度上，这确实是一种进步。但是，假设一个司机住在新泽西州纽沃克的一个偏远地区，她在蒙特克莱尔富人区的一家星巴克工作，每天上班路程有13英里。她的工作时间安排极不规律，有时候

还要负责“关开店”。也就是说，她可能需要在11点关闭咖啡店，开车回到纽沃克，再在凌晨5点前返回。为了每趟节省10分钟并省去在花园州公路上开车需要缴纳的1.50美元高速费，她找了一条近路，这条路的路边布满酒吧和脱衣舞俱乐部。

一家数据驱动的保险公司可能会注意到，深夜在这条路上开车出事故的风险更大。这条路上有不少醉鬼。公平来说，这位咖啡师为了抄近路而不得不和酒吧出来的醉鬼走同一条路确实增加了一点风险——没准就会有个酒鬼撞上她的车。但保险公司的地理跟踪器做了进一步的推断：她不仅与醉鬼走了同一条路，她还可能是其中之一。

就像这样，这些追踪我们并试图分析我们的行为模式的模型通过把我们与别人比较而对我们进行推断，并据此进行风险评估。这一次，它们没有将人们归类为说阿拉伯语或者说乌尔都语的人，住在同一个地区的人，或者挣同样工资的人，而是把人们按照相似的行为模式进行归类。保险公司预测，有类似行为模式的人具有同样程度的风险。可以看到，这又是一个“物以类聚，人以群分”的例子，同样的不公正依然存在。

我和很多司机讨论过他们卡车里的监视器。比起反对数据分析，他们更反对这种监视行为。他们坚持对我说他们不会向监视器投降。他们不想被跟踪，也不希望自己的信息被卖给广告商，或者被交给国家安全局。其中一些人也许最终能成功抵制这种监视，但是保护隐私的成本的确越来越高了。

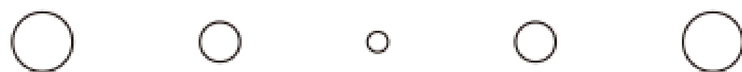
早期，汽车保险公司的追踪系统只是一个可选项。只有那些愿意被追踪的人需要打开安装在车内的监视器。他们得到的奖励是5%~50%的保险折扣，以及将来的更多承诺。（而我们其余的人则要支付更高的保险费来补贴这些折扣。）但随着保险公司获得的信息越来越多，它们的风险预测能力越发强大。这就是数据经济的本质。那些从客户信息中获取最多情报的保险公司成了这个行业中利润率最高

的佼佼者。保险公司能更准确地预测群体的风险性（尽管对个体的推断常常出错）。而且从数据中获利越多，保险公司就会越发推崇数据。

在某一时刻，追踪器很可能会发展成为常态。而那些想要以旧方式买保险的消费者，为了保护仅有的一点隐私，将不得不支付额外的保险费，而且很可能极为昂贵。在数学杀伤性武器泛滥的世界里，隐私将逐渐变为一种只有富人才负担得起的奢侈品。

与此同时，监控将改变保险的本质。传统上，保险业是一个让大多数人满足不幸的少数人的需求的行业。在遥远的过去，我们住在村镇里，当火灾、事故或者疾病发生的时候，各个家庭、宗教群体以及邻里之间会互相照看。而在市场经济，我们把这种关心外包给保险公司，保险公司把一部分钱留给自己，并称之为利润。

由于保险公司对我们的了解越来越多，它们现在已经能够查明哪些人是风险最高的客户，然后，它们要么将这些人的保险费增加到最高，要么在合法的限度内拒绝支付他们的保险赔偿金。这与保险业创建之初帮助社会平衡风险的最初目的相距甚远。在一个每个人都被视为可锁定目标的世界里，我们不再是支付平均费用，而是承担了预期的成本。保险公司非但没有帮助我们平稳渡过生活中的意外时期，反而要求人们提前为还未发生的意外买单。这破坏了保险的本意，对那些只能勉强负担保险的人来说，其遭到的打击尤为严重。



保险公司通过审查我们的生活模式和身体情况，把我们归入一些新的群体分类。这些群体分类的产生不是基于年龄、性别、资产净值或者地理位置等传统的参数，而是基于我们的行为模式，并且分类几乎完全是由机器程序推进的。

要想知道这种分类方式是如何增殖扩散的，我们可以来看一下纽约市一家数据公司“感应网络”（Sense Networks）的案例。十年前，感应网络公司的研究员通过分析手机数据了解使用者的移动轨迹。这份数据是匿名的，由欧洲和美国的电话公司提供。（当然，通过考察某个数据点每天晚上停留的地方对该数据的家庭地址做出推断并关联其个人信息，并不是难事，但是感应网络感兴趣的不是个体情况，而是群体行为模式。）

公司的研究团队用纽约市的手机数据训练其开发的机器学习系统，公司没有向该系统提供过多的额外指导。公司不要求系统筛选出郊区居民或者千禧一代，或者分类出不同的消费群体。公司希望系统自行发现数据中的相似性。许多相似性没什么意义，比如每天在街上度过半天的人，或者每天在外面吃午餐的人。但是如果系统充分研究了数百万个这样的数据点，更有意义的模式就会出现，相关性也会出现，甚至一些许多人永远想不到的关联也会显露出来。

随着时间的推移，感应网络的计算机系统吸收了大量的数据，数据点开始被标记为不同的颜色。有些呈红色，有些呈黄色、蓝色和绿色。不同的群体出现了。

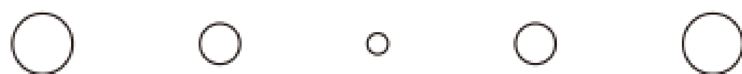
这些群体代表什么？只有系统知道，但是系统不会说话。“我们没必要搞清楚这些人的共同点，”感应网络公司的联合创始人、前首席执行官格雷格·斯奇比斯基表示，“这些群体不符合我们传统的分类标准。”由于不同的群体被标记为不同的颜色，感应网络的研究团队可以分别追踪不同颜色的群体在纽约市内的活动轨迹。在白天，某些地区会被蓝色主宰，在晚上，同一地区又被红色主宰，还有些星星点点的黄色。一个被称为“斯奇比斯基”的群体经常在深夜去某一个特定的地方。那里是个跳舞俱乐部吗？还是一个聚众吸毒的地方？感应网络查看地址，发现那里是个医院。也许该群体经常受伤或者生病，或者他们本人就是医生、护士、急诊医护人员。

感应网络公司2014年被美国电话电报公司（AT&T）旗下的移动广告公司YP收购。所以，目前，这种分类将被用来锁定不同的广告接收群体。但是你可以想象，吸收了大量的不同类型的行为数据的机器学习系统在不久的未来将会把我们归类为数百个具有不同特质的群体，而不再是仅仅将我们归入某一个群体。某些群体会被类似的广告吸引进而消费，某些群体可能有相似的政治倾向或者入狱频率更高，某些群体可能爱吃快餐。

我想说的重点是，海量的行为数据将在未来被直接用于“供养”人工智能系统，而这些系统是我们人眼看不见的。整个过程中，我们几乎无从知道我们属于或者为什么属于某个群体。在机器智能的时代，绝大多数的变量将永远不为人知晓。由于系统不停地对人类进行分类，很多群体每小时、每分钟都在发生变化。毕竟，就算是同一个人，早八点和晚八点的表现也是不同的。

这些自动化程序将日益决定着我们会如何被其他系统对待，后者将决定我们看到的广告，为我们设定商品的价格，为我们预约皮肤科医生，或者为我们规划路线。这些系统运作高效，表面上看只是给出了随机的建议，同时完全不给出任何的解释。没有人能够理解这些系统的内在逻辑或者解释我们接收到这些信息的原因。

如果我们不夺回一定程度的控制权，这些未来的数学杀伤性武器将会成为隐藏在人类社会幕后的控制者。它们将以它们的方式对待我们，而我们却对此毫不知情。



1943年，在第二次世界大战战况最激烈的时段，美国军队和工业急需军人和工人，为此，美国国税局调整了税收法规，为提供医疗保险的雇主免税。这似乎并不是什么大事，显然不能与德国在伏尔加格

勒投降，或是盟军在西西里岛登陆的头条新闻相提并论。当时，只有9%的美国工人拥有作为一种工作福利发放的私人医疗保险。但是，由于新税法的颁布，公司得以通过提供免费的医疗保险吸引员工。在十年内，65%的美国人被纳入其所在公司的保险系统。公司已经在很大程度上控制了我们的资产，而在那十年中，有意或无意地，它们又获取了对我们身体的掌控。

75年后，医疗保险费用转移，如今，全美每年的医疗保险支出达3万亿美元。我们所赚的钱当中差不多有1/5都投入到了庞大的医疗产业。

那些一直以来吝啬地对待自己的员工以降低公司运营成本的雇主，如今找到了降低这些不断增加的成本的新策略，并美其名曰“健康计划”。这一策略涉及日益加强的监视带来的大量数据，包括来自物联网的大量信息——Fitbits计步器、苹果智能手表以及其他的一些感应设备，这些设备将实时收集关于我们身体状况的数据。

我们在本书已经看过很多次了，这种理念的出发点是好的。事实上，政府是鼓励这种做法的。《平价医疗法案》（或“奥巴马医改”）鼓励各公司呼吁员工参加健康计划，甚至提出可以用物质激励积极参加健康计划的员工。根据法案，雇主现在可以为员工提供此项物质奖励，并且可以从政府处报销高达保费50%的保险赔偿金。美国兰德公司的一项调查显示，如今超过半数的雇员人数在50人以上的组织机构已经实行了健康计划，并且每周都有更多的组织机构加入此行列。

实行健康计划的理由很充分。如果这些计划奏效（我们将会看到，这是个大胆的“假设”），最大的受益人将会是员工和他们的家人。而如果健康计划的确帮助员工预防了心脏病或糖尿病，雇主也将获益。一个公司的员工被送往急诊室的次数越少，该公司的全体员工的保险风险也就越小，其保费就越低。所以，如果我们能暂时忽略这项计划对员工隐私的侵犯，健康计划似乎可以被视为一件双赢之事。

然而，对隐私的侵犯不容忽视，也不会因为我们的希望而消失。就拿阿隆·艾布拉姆斯的遭遇为例。他是弗吉尼亚州华盛顿与李大学的一位数学教授。他参保的公司是安森保险，该保险公司也推出了一项健康计划。要达到该计划的要求，他必须累积3250分“健康积分”。每天登录官网可获得1积分，每年看病一次以及参加校园健康体检一次可分别获得1000积分。另外，填写完成一份健康调查问卷也可以获得相应积分，在调查问卷中，他要给自己制订每月的健康计划，如果完成了这些计划，他还可以获得额外的积分。如果他选择不参加这项健康计划，则必须支付每月50美元的额外保费。

艾布拉姆斯受聘在大学教数学。和数以百万计的美国人一样，现在，他工作的一部分内容就是遵守一系列的健康规定，并和他的雇主以及第三方——推出健康计划的公司——分享他的身体数据。他对此感到很厌恶，并且他预感总有一天大学将扩大它的监视范围。他表示：“想到根据我自己记录并上报的移动轨迹数据，有人能再现我一天的活动，我就感觉毛骨悚然。”

我的恐惧比他更进了一步。一旦公司积累了大量有关员工健康的数据，还有什么能阻止它们创建健康评估模型并将其运用于筛选求职者呢？很多被收集的可以作为替代变量的数据，不论是每天的步数还是睡眠模式，都不受法律的保护，所以这种行为理论上来讲是完全合法的，并且具有重大意义。正如我们之前所见到的那样，基于信用评级以及性格测试，公司会按照一定的标准拒绝一部分求职者。而由健康评估模型给出的健康指数则代表的是筛选机制自然而然且令人恐惧的下一发展阶段。

一些公司已经给员工设定了较高的健康标准，并且对没有达到要求的员工予以处罚。米其林公司就给员工设定了一个健康标准，其评估范围包括血压、血糖、胆固醇、甘油三酯以及腰围等各方面。有三

项及以上指标未能达标的员工，每年需要额外支付1000美元的医疗保险费用。

全美药品连锁公司CVS在2013年宣布了一项内部规定，要求员工报告他们的体脂、血糖、血压和胆固醇的水平，如不上报，则需每年支付600美元。

对于CVS的这一举措，Bitch Media的一位专栏作家，阿丽萨·弗莱克表达了她的愤怒："大家注意了：如果你多年来想方设法试图保持体型却一直没能成功，现在你可以放弃你正在尝试的方法了，因为CVS替我们找到了终极答案。原来，不论你之前一直在尝试什么愚蠢的减肥方式，你没能成功只是因为你没有恰当的动力。而现在，正如CVS所规定的，终极减肥法已经出现了，就叫作羞辱法或者以肥胖为耻法——要么让某人告诉你你已经超重了，要么就付罚金。"

体重问题的核心是一项不足为信的数据——身体质量指数（BMI）。该指数建立在两个世纪前一位比利时的数学家兰伯特·阿道夫·雅克·凯特莱提出的一个公式之上，而此人对于健康或者人体几乎一无所知。他只是想用一个简单的公式来评估大量人口中的肥胖情况。他是以他所谓的"一般人"的指标为基础创造出这个公式的。

数学家、科学作家基思·德夫林写道："那是一个有用的概念。但如果你试图把它运用于任何一个个体，你就会觉得这跟说一个人有2.4个孩子一样荒谬。'一般人'的水平可以用于衡量整个人口但往往无法用于评估个人。"德夫林还指出，身体质量指数通过给出一个数值评分，给这个万灵药概念增添了科学的权威性。

身体质量指数的计算方式是用一个人的体重公斤数除以身高厘米数的平方。它是一个人身体健康情况的粗略的数值替代变量。根据这种计算方法，女性更有可能超重（毕竟"一般人"并不真的存在）。而且，因为脂肪比肌肉轻很多，健美的运动员通常会有极高的身体质量

指数。在一个由身体质量指数统治的世界中，篮球名将詹姆斯·勒布朗也可以算得上超重。当身体质量指数和经济上的“软硬兼施”策略相互捆绑时，一大批员工因为他们的身材而受到了处罚，尤其是黑人女性，因为她们的身体质量指数通常都偏高。

但是，健康计划的倡导者可能会说，帮助人们控制自己的体重并预防一些相关的疾病，难道不是一件好事吗？但核心的问题在于加入此类健康计划是出于员工自愿还是公司强制要求。如果公司制订的都是可自由加入的自愿的健康计划，那么没有人会有理由反对。（而且选择加入该计划的员工可能确实会有所收获，尽管他们即便不参加该计划也能很好地达成目标。）但是，用一个有漏洞的数据，如身体质量指数，去衡量员工，如拒绝就让他们支付罚金，并强迫员工按照一个所谓的理想指标去改造他们的身材，这种做法侵犯了员工的人身自由，也给了公司雇主惩罚其不待见的员工，同时处以他们罚款的借口。

所有这些做法都是借健康之名义推行的。与此同时，价值60亿美元的健康产业正在毫无依据地大肆宣扬它们的成功。合资健康公司Kinema Fitness的董事长约书亚·洛夫这样写道：“有事实为证：健康的员工工作更加努力，更加快乐，更乐于助人且工作更加高效；反之，不健康的员工则普遍较为懒散，时常陷入过度疲劳且闷闷不乐的状态，他们的工作状态就像是他们的不良生活方式的一种症状。”

自然，洛夫并没有给这些笼统的断言以例证。而且，即使他所说的是真的，也没有足够的证据证明强制的健康计划确实能使员工更加健康。根据加利福尼亚健康福利审查项目小组的一项研究报告，公司推行的健康计划未能有效地将参与其中的员工的平均血压、血糖或胆固醇值维持在正常范围内。即使是有人在加入了某个计划之后成功减重，他们往往也会胖回来。（健康计划确实起到积极作用的唯一方面是戒烟。）

而且，尽管个别的成功案例被广泛宣扬，健康计划实际上并没有广泛降低人们的医疗费用。加州大学洛杉矶分校的一位法律教授吉尔·霍维兹在2013年组织了一项研究，研究报告揭示了这项健康计划运动的经济学基础。研究者进行的随机性实验对吸烟者和肥胖的员工会比别人承担更多的医药费这一定论提出了质疑。尽管这些人确实更可能面临健康问题，但这些疾病往往发作于患者晚年，那时他们早已退出了公司的健康计划，被纳入了老年人医疗保险。事实上，健康计划节省的最多的钱就是雇主为员工缴纳的保险赔偿金。换言之，正如工作时间调度算法那样，此类模型反而为企业提供了一种新的克扣员工薪水的手段。

尽管我对健康计划有所质疑，但它目前还不完全算是数学杀伤性武器。它是自然传播的，且已经渗透了数以百万计的员工的生活，尽管可能给员工造成了经济损失，但它仍是公开透明的，并且除了身体质量指数得分外，计划的其他部分并不是基于数学算法建立的。目前，它仅仅是流传甚广的、披着花哨的健康言论外衣的克扣薪资的工具。

公司的雇主们已经过分沉溺于我们的各种数据。正如我们所见，他们频繁地使用这些数据，去衡量我们是否是有潜力的雇员和工人。他们试图借助数据勾勒出我们的思想和我们的朋友圈，并预测我们的生产力。由于员工的医疗保健是公司的一笔主要支出，而员工的保险款项已经被写进法律明文之中了，因此，大范围扩大对员工的身体健康状况的监视对他们而言再自然不过了。如果公司最终编写出了自己的健康和生产力评估模型，那么这将必然演变成一个完全成熟的数学杀伤性武器。

第十章
正面的力量

微目标的出发点



现在你已经知道，我对各种各样的数学杀伤性武器感到愤怒。所以，让我们想象一下，如果我现在决定发起一场运动，要求政府加大对数学杀伤性武器的监管力度，并且在我的脸书页面上发布一份请愿书，那么我的哪些朋友会在他们的脸书信息流中看到这份请愿书呢？

我不知道。点击发送后，这份请愿书就属于脸书了，而社交网络的算法会对如何最大限度地利用这份请愿书做出判断。算法会计算出请愿书吸引我每一个朋友点击的概率。算法知道我的某些朋友经常签署请愿书，而且很可能会把这些请愿书分享到他们自己的脸书页面上，我的另一些朋友则往往对此类信息一扫而过。与此同时，我的一部分朋友更常关注我，会经常点开我发布的文章。脸书的算法在决定谁能看到我的请愿书时会把所有这些都考虑在内。结果是，对于我的

很多朋友而言，我的请愿书只会出现在他们信息流的底部，他们永远也不会看到。

这就是拥有15亿用户的互联网巨头、上市公司所做的事情。脸书看起来就像是一个现代化的城镇广场，但它可以根据自身的利益决定我们在其平台上看到和学习到的内容。在我写作本书的时候，约有2/3的美国成年人在脸书上建立了自己的个人页面。他们每天在该网站上的停留时间平均为39分钟，比其每天面对面的社交活动平均时间只少了4分钟。根据皮尤研究中心的报告，近一半的使用者将脸书作为其发布消息的主要渠道，这就引出了一个问题：通过调整算法，规定用户能看到的信息，脸书是不是就能钻政治体系的漏洞了？

脸书的内部研究人员一直在研究这一问题。在2010年和2012年的美国总统大选中，脸书的员工做了一些实验以测试他们开发的一种被称为“选民扩音器”的工具。该工具的核心概念是鼓励人们广泛和他人分享他们的投票经历。这似乎非常合理。通过鼓励用户以“我已投票”为标签推送信息，脸书鼓励了6100多万美国人履行了他们的公民义务，并让他们的声音被更多的人听到。在此基础之上，通过推送投票的相关信息，脸书也激起了好友间的投票压力。研究表明，就对促进个体投票的有效性而言，履行一项公民义务所带来的满足感没有因为没投票而被朋友、邻居说三道四强。

与此同时，脸书的内部研究员也在探究不同类型的信息推送会如何影响人们的投票行为。从未有研究人员从事过如此大规模的人类实验。脸书能够在几小时之内收集数千万条甚至更多的数据，用以衡量每一个人对其投票行为的描述用语和互相分享过的链接带来的影响。然后，脸书就可以利用所得结论影响人们的行为，在这一政治选举的案例中，脸书影响的就是用户的投票行为。

这种权力是巨大的，而且拥有并行使这种权力的并不只有脸书一家公司。其他的上市公司，如谷歌、苹果、微软、亚马逊，以及威瑞

森和美国电话电报公司等手机供应商也掌握着大量关于用户行为的数据以及引导我们做出特定选择的无数方式。

正如我们所见，这些公司往往专注于赚钱。但是，这些公司的利润与政府政策紧密相关。政府监管它们，或者不监管它们，批准或阻止企业的合并和收购，制定税收政策（而对放置于海外避税天堂的数十亿美元往往视而不见），都会对这些公司的收益产生影响。这就是为什么科技公司，像美国的其他公司一样，不断地向华盛顿输送它们的说客，并在暗中向政治体系注入数亿美元的资金。现在，它们已掌握了必要的手段和资金用于调整我们的政治行为，从而塑造美国政府的形象。

脸书发起的这项运动是出于一种有意义的、看似单纯的目的：鼓励人们投票。脸书成功了。通过比较投票记录，研究人员估算这项运动共促进了34万本来没有投票意图的人参加投票，这一增量足以改变整个州甚至全国的选举结果。2000年，乔治·W.布什以537票的微弱优势赢得了佛罗里达州的选举。对比来看，单单一项脸书的选举日活动就能打破国会的席位分布，甚至决定谁将最终当选总统。

脸书的影响力不仅来自它的大用户量，还来自它拥有的操纵用户影响他们的朋友的能力。在参加此次选举日大型实验的6100万用户中，绝大多数人都在他们的信息中看到了一条鼓励他们去投票的信息推送。这条信息包含一组照片：用户的6个由网站随机选择的点击过“我已投票”标签的脸书好友。研究人员还设置了两组控制组，每组大约有60万人。一组人看到了那条鼓励他们去投票的信息推送，但其中没有他们的朋友的照片。另一组人则什么相关的信息推送也没收到。

通过改变信息推送的方式，脸书研究了我们的朋友的行为对我们自身的影响。人们会鼓励自己的朋友投票吗？他们的朋友会受此影响吗？根据研究人员的统计，朋友参与与否完全改变了我们的行为。当

带有“我已投票”标签的信息推送是来自他们的朋友时，人们更容易注意到这些信息，并且更有可能分享这些信息。大约有20%的人在看到他们的朋友发布了带有“我已投票”标签的信息之后也加入了该活动。而那些没有从朋友那里得到相关信息分享的人中，只有18%的人参与了投票活动。的确，我们不能确定所有参与该活动的人都真的去投票了，或者那些没有参与活动的人就一定没投票。尽管如此，对于6100万的潜在网络选民而言，两个百分点的差距也可能带来巨大的影响。

两年后，脸书的算法发展得更加先进了。在奥巴马和米特·罗姆尼竞选总统的前三个月，脸书的数据科学家所罗门·梅辛改变了约200万政界人士的脸书页面的信息推送算法。他们会接收到更多的重大新闻，而不再被小猫视频、毕业典礼邀请或者迪士尼乐园的照片等占据信息流。如果他们的朋友分享了一条新闻报道，这条新闻报道会很快出现在这些人的信息流的优先位置。

所罗门·梅辛想知道，从朋友处收到更多的严肃新闻信息能否改变一个人的政治行为。大选之后，所罗门·梅辛展开调查。调查结果表明，被改变信息推送算法的用户组的投票率由64%增加到了67%。脸书的计算社会科学家拉达·阿达米克说：“你的朋友在分享新闻的时候，有趣的事情也同时发生了。”当然，新闻实际上并不是真的由朋友分享的，而是由脸书在后台推送的。你可能会说，新闻报纸也有同样恒久的影响力。主编们会挑选头版新闻并决定怎样描绘整个故事。他们会决定是站在战乱中的巴基斯坦人的视角还是站在哀号的以色列人民的视角来叙述这个新闻故事，或者决定是特写救助了一个婴儿的警察还是特写打击了一个反抗分子的警察。毫无疑问，这些选择不仅会影响公众的舆论，也会影响选举结果。电视新闻也是一样的道理。但是，《纽约时报》或者CNN电视台报道的新闻是所有人都看得到的，主编们所做的报道决定显而易见，公开且透明。而且事后人们还可以（经常是在脸书上）讨论主编们的决定是否正确。

而脸书则更像是绿野仙踪：我们看不到有所谓的“主编”存在。当我们浏览网站的时候，我们看到的是来自朋友的推送。从表面上看，机器只是一个中立的中间人。很多人现在仍然相信事实就是如此。2013年，伊利诺伊大学的研究者凯瑞·卡拉哈里奥调查了脸书的算法，她发现，62%的用户不知道该公司会对信息推送进行暗箱操作。他们认为，系统只是不断把他们发布的所有内容分享给他们的朋友。

脸书对大众政治生活的影响远不止于改变信息推送算法和动员投票活动这么简单。2012年，脸书的研究小组对68万脸书用户展开了一项实验，研究他们实时更新的信息推送能否影响他们的情绪。毕竟，已有实验研究表明，人们的情绪极易受到他人情绪的感染。简单来说，周围人的坏情绪很可能也会让你的心情变差。那么，这种情绪感染也适用于网络上的信息传播吗?

脸书利用语言识别软件将积极信息和消极信息进行分类，然后对于一半的用户减少其信息流中消极信息推送的数量，对于另一半用户则减少其信息流中积极信息推送的数量。当研究用户的后续信息发布行为时，他们发现，有证据表明调整过的信息推送算法确实改变了用户的情绪。在信息流中看到更少积极信息推送的用户会发布更多负面的帖子，反之亦然。

研究者得出结论：“人的情绪状态能够传染给其他人……人们会因此无意识地与周围的人拥有同样的情绪状态。”换句话说，脸书的算法能够影响数百万人的感觉，而与此同时，用户们却被蒙在鼓里。试想一下，如果脸书在选举日操控人们的情绪又会如何?

我没有理由认为脸书的社会科学家正在想方设法试图操控整个政治系统。他们中的大多数人都只是在一个以往仅存于幻想中的平台上做了20多年研究的严谨学者。但是，他们展示出了脸书在我们能学到什么、有怎样的感觉以及是否投票等方面的巨大影响力。脸书这个平

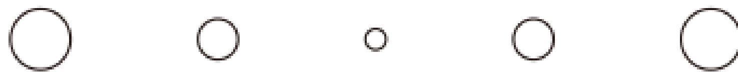
台巨大、强力，而且不透明，其算法不为我们所见，我们只能看到研究者选择公布的实验结果。

谷歌也大致如此。谷歌开发搜索算法的主要目的也是增加收入。而经过谷歌筛选的搜索结果会对人们了解了什么信息以及他们如何投票产生巨大影响。罗伯特·伊比斯坦和罗纳德·E.罗伯森两位研究者让美国和印度两国中犹豫不决的选民借助搜索引擎了解即将到来的选举的竞选者信息。他们对搜索结果进行了设定，使其偏向于其中一个党派。两位研究者称有偏向的搜索结果改变了20%的选民的投票选择。

这种巨大的影响力部分来源于大多数人对搜索引擎的信任。据皮尤研究中心的一份报告，约73%的美国人相信搜索给出的结果既精确又公正。所以，如果像谷歌这样的公司刻意使搜索结果偏向大选中的某一个党派，那么它们就是在拿公司的名誉做赌注，而且极有可能遭到监管机关的处罚。

但同样地，又有谁知道它们这样做了呢？我们对这些互联网巨头的了解主要来自它们公开的极少部分研究报告。它们的算法是重要的商业秘密，这给予了它们暗箱操作的权限。

我暂且不将脸书或者谷歌的算法视为政治领域的数学杀伤性武器，因为没有证据表明这些公司正在利用它们的算法伤害大众。但这些算法被滥用的可能性是巨大的，而且这些事都将发生在程序代码之中和防火墙之后。正如我们即将看到的，这些技术可以将我们每个人都置于我们政治上的舒适角落中不再出来。



2012年春末，前马萨诸塞州州长米特·罗姆尼被提名为共和党的总统候选人。罗姆尼的下一步就是建立他的竞选基金，以便与时任总统奥巴马进行角逐。因此，同年的5月17日，他前往佛罗里达州的波卡拉

顿，参加一个由私人股本投资者马克·莱德在其宫殿似的家中举办的筹款活动。莱德已经向“支持罗姆尼的超级政治行动委员会，重塑我们的未来”基金投入了22.5万美元，又为罗姆尼的竞选委员会提供了6.3万美元的资金，他召集了一群富有的朋友与这位总统候选人见面，他们中的大多数从事的都是金融业和房地产行业。自然，他们有一场宴席要办。

罗姆尼可以放心地认为，他即将进入的是这样一个封闭的环境，其中在场的大多数人都和马克·莱德一样会选择支持他。如果这是一次电视讲话，罗姆尼会非常小心地避免激怒他潜在的共和党选民，这些人包括福音派基督徒、华尔街金融家、古巴裔美国人和中产家庭的母亲。试图取悦所有人是大多数政治演讲都很枯燥乏味的原因之一（罗姆尼的演讲尤其乏味，连他的支持者都对此颇有微词）。但是，在马克·莱德家的这次内部人士聚会上，一个规模虽小却极富影响力的团体可能能够接近真正的米特·罗姆尼，听到这位总统候选人真正的、未经美化的观点。他们已经给了他大笔的捐款，现在，他们至少能从这位总统候选人的嘴里听到一点儿真话了。

包围在他认为是支持他的、志同道合的人之中，罗姆尼开始畅所欲言。他指出，47%的人口都是“索取者”，依靠政府的慷慨施舍过活。他说，这些人永远不会投他的票，这使得取得另外53%的人的支持变得尤为重要。但事实证明，罗姆尼的判断失误了。在这场内部人士的聚会的角落，为这些富人提供服务的都是局外人。和所有其他人一样，他们也有带有拍摄功能的手机。罗姆尼的轻蔑言论被宴会中的一名酒保忠实记录下来，视频在网络上疯传。这次失言很可能让罗姆尼就此丧失入主白宫的机会。

要想这次聚会取得成功，罗姆尼既需要准确定位其支持者，也需要信息的严格保密。他的确希望成为马克·莱德和其富人朋友们心中的理想候选人。而且，他也曾相信马克·莱德的家对于一个总统候选人来

说是安全的。在理想的情况下，政客们会来往于数不清的诸如此类的由不同的潜在支持者建构的安全区域，这样他们就能在面对不同的群体时斟酌不同的措辞，而不让外部人士听到。一个候选人需要让自己面面俱到，只向特定的选民群体展示他们想看到的那一面。

这种政客的两面性或“多面性”在政治领域已经不新奇了。政客们可以在密尔沃基市吃波兰香肠、在布鲁克林引用圣经，同时在艾奥瓦州许诺保护玉米乙醇燃料，他们一直以来都试图向不同的群体展示不同的侧面。但罗姆尼已经知道了，如果做得太过分，视频摄像头可以轻易毁掉他的形象。

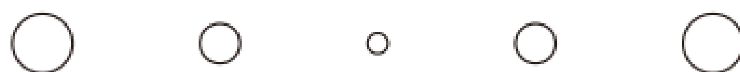
但是如今，现代消费者市场为政客们提供了争取特定选民的新途径，政客们又可以说出选民想听的话了。而一旦他们这样做了，那些选民就有可能欣然接受这种信息，因为这些信息与他们本来的观点相符。这种现象被一位现象心理学家称为证实性偏见。这也是聚会中的受邀宾客没有人去质疑罗姆尼的主张的原因之一。罗姆尼认为几乎半数的选民都渴求得到政府的援助，这一主张正与在场宾客既有的观点相一致。

这种政治与面向消费者的市场推广的结合已经发展了半个世纪，接近、说服政治名流，给潜在的赞助人挨个打电话等传统方式已经逐步被市场营销科学取代。在1968年理查德·尼克松的总统竞选活动结束之后，记者乔·麦金尼斯在其《总统的营销》（*The Selling of the President*）一书中就向读者介绍了在政治领域中把总统候选人当作一件商品进行营销的竞选团队。通过将选民按一定特征进行分类，尼克松得以在竞选活动中针对不同地区、具有不同人口特征的选民群体打磨自己的措辞。

然而，随着时间的推移，政客们开始想要一种更加精确的营销方式，一种在理想情况下可以接触到每一位投票者，给他们以更具针对性的诱惑的方式。这一需求催生了直邮竞选宣传。借鉴信用卡行业的

策略，政治营销人员创立了关于潜在选民的庞大数据库，并根据他们的价值观和他们的群体特征把他们归入不同的小组。这就导致有史以来首次出现了某人和他的隔壁邻居收到来自同一个候选人的不同内容的宣传邮件或宣传册的情况，可能一封是承诺建立野地保护区，另一封则强调了法律和秩序。

直邮竞选宣传对于每一个选民的针对性投放，受益于大数据技术和市场营销的结合。现在，竞选团队可以精准锁定每一个微型市民群体，争取他们的选票和赞助，用精心打磨过的针对性的宣传信息来吸引他们的注意，而这些信息是任何其他群体都看不到的。这些宣传信息也可能以脸书上的一则标语或是一封筹集资金的电子邮件等形式出现。每条宣传信息都可以让候选人悄无声息地展现他们的多面性，而谁也说不准他们在真正当选后会展现其中的哪一面。



2011年7月，在奥巴马总统开始其连任竞选的一年多之前，一位名叫伊德·加尼的数据科学家在领英上发布了一则招聘信息：

> 招聘有志于创造性工作的分析学专家。奥巴马总统连任竞选工作小组正在组建一个分析学专家团队，负责高强度、大规模的数据挖掘工作。
>
> 现有若干岗位需求不同经验背景的人才：统计科学家、机器学习专家、数据挖掘专家、文本分析专家以及预测分析专家。该团队的工作是处理大量的数据，为竞选策略提供指导。

加尼毕业于卡耐基–梅隆大学，是一位计算机科学家，他领导了奥巴马竞选工作小组的数据团队。他之前在芝加哥的埃森哲咨询公司工作，开发面向消费者的大数据应用程序，他相信他的技术也可以应用

于政治活动之中。奥巴马竞选小组的目标是创建以意向相似为标准的选民分组，每个分组中选民的价值观和关心的议题是一致的，就好比罗姆尼在迈克·莱德家接触的赞助者一样——但这次避开了服务生的耳目。然后他们就可以面向这些选民，发送一些有针对性的、能够促使选民做出某种决策，以达到他们既定目标的宣传信息，比如促使投票、组织活动和筹集资金等方面的信息。

加尼在埃森哲公司做过一个给超市购物者建模的项目。一个大型食品商为埃森哲的一个团队提供了一个庞大的、匿名的消费者购买行为数据库。加尼的想法是深入挖掘该数据库，研究每个消费者的购买习惯，继而把消费者归入数百个不同的购买组别里。冲动购物者会在收银台买糖，注重养生的人愿意花3倍的价格去买有机甘蓝。这些是很明显的购物者类别。但还有一些更令人意想不到的类别。比如说，加尼和他的团队发现了钟爱某个特定品牌的购物者和因为一个小小的折扣就换个品牌购买的购物者这两个对立的分组。团队的终极目标就是为每个购物者制订一个专属的购物计划，以指导他们的购物过程，引导他们去最容易产生购买意向的食品区消费。

对于埃森哲公司的客户来说，不幸的是，这个终极愿景取决于携带计算机程序的购物车的发明，这种购物车至少在目前还没有大规模地流行起来，而且可能永远不会。不过，尽管在超市购物者建模这个项目上碰了壁，但加尼的技术被完美地应用到了政治领域。比如说，那些因为想省几分钱就换个品牌购买的善变的购物者，就特别像摇摆不定的选民。估算出每个超市购物者从选择一个品牌的番茄酱或咖啡转向选择另一个更划算的品牌超市所需给出的折扣是可能的。在得出这个计算结果之后，超市就可以挑出15%的最有可能摇摆的顾客，给他们发放优惠券。智能的目标锁定是很有必要的。超市必然不想给那些乐意购买某一品牌的正价商品的顾客发放另一品牌的同类商品的优惠券，那无异于烧钱。[注]

那么，类似的算法适用于摇摆不定的选民吗？在得到了包含消费行为、人口特征以及投票行为信息的庞大数据库之后，加尼和他的团队开始着手研究这一问题。然而，超市模型与选民模型之间有一个关键的不同。在超市模型中，所有可用数据都与购物密切相关。他们通过研究人们的购物模式，来预测（和影响）人们的购物行为。但在政治选举方面，密切相关的可用数据则少之又少。要建立这个模型，数据团队需要找到替代变量，而这需要建立在深入调查的基础上。

于是，他们首先面向几千人进行了深度采访。这些人被归入不同的组别。一些人关心的是教育或者同性恋者的权利，另一些人担心的是社会安全问题或者淡水含水层的水力压裂。一些人无条件地支持奥巴马总统，另一些人则持中立态度。很多人很欣赏奥巴马总统但是往往不会抽时间去给他投票或说服别人去投票。还有一些人——这些人是最关键的——愿意给奥巴马的竞选活动提供资助。

一旦加尼的数据团队搞懂了最后这一小部分选民样本的需求、顾虑，以及怎样影响他们的实际行动，下一个挑战就是找到数百万类似的选民和赞助者。这需要他们仔细研究数据并总结他们所采访的选民的行为模式和人口统计资料，并为他们创建数字档案。然后就是搜索全美的数据库，找到有相似档案的人并把他们分入同一个组别。

然后竞选工作小组就可以针对每个组别的选民发送有针对性的宣传信息了——可能是在脸书上，也可能是在他们经常访问的媒体网站上，看一看他们的反应是否与预期相符。他们采用A/B测试法，就是谷歌用来测试标题字颜色采用哪种色度的蓝色能获得更多点击量时采用的方法。在尝试了各种信息投放方法后，他们发现，那些标题只写了“Hey！”的邮件会打扰到人，但也更容易吸引人点击查看，从而换取更多的赞助资金。经过成千上万次的测试和调整，竞选工作小组终于找出了所有的目标选民，其中包括非常重要的1500万摇摆不定的选民。

每个竞选工作小组都会在此过程中创建美国的选民档案。每个档案包括很多项评分，这些评分不仅被用于衡量他们作为一个潜在的选民、志愿者和投资人的价值，同时也被用于反映他们对不同议题的关心程度和立场。一个投票者可能在环境问题上得到了很高的评分，但在国际安全和国际贸易上则得到了很低的评分。这些政治档案和亚马逊、网飞等互联网公司为管理数千万用户而创建的那些档案非常类似。那些公司的数据分析模型几乎是在不间断地进行成本/效益分析，以从每个顾客身上获取收益的最大化。

四年后，希拉里·克林顿的竞选工作小组沿用了奥巴马团队创建的这套方法论。希拉里竞选团队与数据挖掘公司the Groundwork签约合作。该公司由谷歌公司的董事长埃里克·施密特出资创建，由2012年奥巴马竞选团队的首席技术官迈克尔·斯拉比负责运营管理。根据技术平台Quartz的一篇报道文章，希拉里竞选团队的目标是打造一个数据系统，然后据此创建一个模型，即类似于赛富时（Salesforce）这类公司开发的用以管理数百万顾客的模型的政治版本。

你可以想象，该模型对于实时更新的相关数据有着极大的需求。而其中一部分收集信息的方法极为恶劣，严重干扰了人们的正常生活。2015年年末，据《英国卫报》报道，政治数据挖掘公司剑桥分析公司聘请英国的学者制作美国选民的脸书档案，档案包含所有的人口统计学信息和每个用户的“点赞”记录。他们利用这些信息分析了超过4000多万选民的心理状态，并按照五大人格特征——开放性、严谨性、外向性、宜人性和神经质——进行了分组。和泰德·克鲁兹的竞选工作小组合作的团队利用这项研究结果为不同类型的选民设计电视广告，把广告安排在选民们最有可能看到的时间和地点上。比如，2015年5月，共和党犹太人联盟在拉斯维加斯的威尼斯人酒店集会时，克鲁兹的竞选团队发布了一系列的网络竞选广告，这些广告只在酒店大楼内可见，广告重点强调了克鲁兹对以色列和以色列安全问题做出的贡献。

在这里我要说的是，不是所有的针对性的广告投放都能奏效，其中一些和骗人的万灵药一样无效。毕竟，这些数据挖掘公司面对的潜在客户是坐拥数百万美元资金的竞选活动小组和政治行动小组。为了争取这些客户，这些公司必然会承诺自己拥有极为宝贵的数据库和精确的选民定位技术，但这样的承诺大部分都是言过其实的。也就是说，政客们不仅是虚假承诺的供应者，也是虚假承诺的购买者（而且往往是高价购买）。即便如此，仍有一些方法回报颇丰，奥巴马的竞选团队已经证明了这一点。所以，在这个行业，不管是严肃的数据科学家还是无良的数据商贩，都将目标对准了选民。

然而，政治领域的数据挖掘公司还面对着特殊的限制，它们的工作因此变得更加复杂。比如说，每个潜在选民的价值上升或下降有时取决于其所属州的立场。摇摆州（比如佛罗里达州、俄亥俄州、内华达州）里的摇摆选民就有很高的价值。而如果民意调查显示某个州显著倾向于民主党或者共和党，那么该州选民的价值就会急剧下降，该州的宣传预算很快就会被投入于拉拢有价值上升空间的其他州的选民。

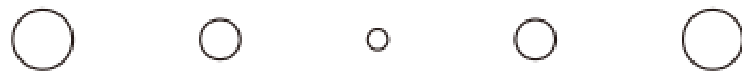
从这个意义上讲，我们可以把参与选举的大众想象成金融市场。随着新信息的涌入，选民的价值会有所上升或下降，相应的投资也随之波动。在这些新兴的政治市场，我们每个人都代表着一只价格持续上下浮动的股票。而每个竞选小组必须决定是否以及以何种方式买进我们。如果我们值得买进，那么他们不仅要决定以什么内容的宣传信息吸引我们，还要决定投放多少宣传信息以及采用何种投放方式。

类似的更宏观的算法已经盛行了数十年，比如，竞选小组会规划他们的电视广告投入。当民意调查的结果发生变化时，他们可能会减少在匹兹堡的广告投入，转而将更多的预算投向坦帕或者拉斯维加斯。但随着微目标锁定技术的发展，竞选小组的关注焦点从地区转移

到了个人。更重要的是，每个选民看到的都是政客专为自己打造的那一面。

竞选小组利用类似的分析方法来锁定每一个潜在的赞助人，然后最大限度地调动他们。而由于每一位赞助人也都打着自己的算盘，这件事情因此变得复杂起来。他们每个人都想把自己的利益最大化。潜在的赞助人知道如果自己轻易地把全部资助额度都交给了竞选团队，那么他们就会被标记为“充分利用”，从而不再对相应的候选人有影响力。但是，如果一点儿钱也不给的话，他们就会被排除在内部人士的圈子之外。所以，很多人都会根据自己是否认同候选人传递的信息而一点一点地给钱。他们认为，对付一个政客就像用食物奖励的方式训练一只狗一样。“超级政治行动委员会”（Super PACS）的赞助人通过这种赞助方式收到了格外好的效果，该委员会不对政治捐款进行限制。

当然，竞选团队也非常清楚潜在赞助人的策略。其应对策略是利用微目标锁定技术给每一位赞助人发送他们想听到的信息，利用这些信息从他们的账户中“套出”更多的钱。可以想见，每一位赞助人收到的信息都不一样。



这些策略的运用并不局限于竞选活动，我们的生活也深受影响，因为说客和利益集团正在利用这种微目标锁定技术开展他们的肮脏交易。2015年，反堕胎组织“医学进步中心”发布了一系列关于计划生育组织旗下的一家诊所如何对待流产胎儿的视频。视频显示，计划生育诊所的医生会把流产胎儿的器官出售给研究机构。视频激起民众的抗议浪潮，共和党也提案撤销这个组织的资助资金。

后来有调查者发现，该视频存在伪造痕迹，视频中所谓的流产胎儿其实是宾夕法尼亚州一个乡村孕妇产下的死胎的照片。计划生育组织未曾贩卖过任何胎儿器官。医学进步中心不得不承认该视频包含错误的信息，该事件削弱了医学进步中心对大众的号召力。但是，尽管如此，反堕胎积极分子仍然可以继续利用微目标锁定技术挖掘该视频的受众，以此募集资金对抗计划生育组织。

上述事件有幸进入了公众视野，而成百上千个类似的行动则隐藏在幕后，将目标对准个体选民。这些在暗地里开展的活动具有同样的欺骗性，甚至还要更不负责任，并且会公开传播政治家在公开演讲时不会明示的隐晦思想。北卡罗来纳大学的科技社会学家泽伊内普·图菲克希教授表示，这些组织精准定位易受影响的选民，针对他们制作引起其恐慌的广告，用他们孩子的安全或者非法移民的兴起恐吓他们。同时，他们还可以确保对此不感兴趣（甚至反感）的选民看不到这些广告。

成功的微目标锁定行动可以部分解释，为什么在2015年的一项调查中，超过43%的共和党人仍然相信奥巴马总统是穆斯林这样的谎言，还有20%的美国人相信他不是在美国出生的，所以让他当选总统是不合法的。（民主党人也在利用微目标锁定技术大肆传播误导信息，但是没有什么竞选活动有像反奥巴马竞选活动那样大的规模。）

尽管微目标锁定技术发展迅速，政治竞选活动仍然把平均75%的宣传投入放在电视上。你可能会认为两者效果差不多，事实确实如此。电视的信息传播面更广，更具公信力，而微目标锁定行动更多的是暗箱操作。但电视广告也开始转向个性化的方向发展了。新兴广告公司，如纽约的Simulmedia，根据电视观众的不同观看行为对他们进行分组，这样广告商就可以锁定相应的目标群体，如狩猎者群体、和平主义者群体、爱买大型越野车的群体投放相应的广告。当电视及其

他媒体也开始为它们的受众建立数字化档案之时，政治微目标锁定的效果也就得到了进一步的增强。

这样一来，要想了解我们的邻居看到了什么样的政治信息就更难了，同时，我们也就更难理解为什么他们会狂热地相信他们所做的事情是正确的。甚至连爱探听消息的记者也很难追踪到这些信息。访问候选人的网页也没有看上去那么简单直白，因为竞选团队也自动地给每个访客建立了数字档案并对他们进行了分类和定位，对各种细节信息进行了评估，如所在地区、点开的网页链接，甚至他们查看的照片。在注册时填入虚假信息也是无用功，因为系统会深度收集每个真实选民的资料，包括他们的购买记录、家庭地址、电话号码、投票记录，甚至社会保障号码和脸书档案。为了向系统证明你创建的虚假档案是真实的，你需要提供大量的具体的数据资料，这意味着巨大的工作量（在最坏的情况下，你所创建的虚假档案甚至会让你卷入欺诈案件）。

这些竞选活动中的暗箱操作带来了危险的不安定因素。这些政治营销者拿到了我们的一手资料，向我们投放一点点信息，然后测试我们的反应。而我们对我们邻居所接收到的信息一无所知。这就像是商务谈判代表使用的一种惯用策略。他们区别对待谈判双方，让一方无从知晓另一方听到的信息。这种信息的不对称可以防止对立两方的联合，而这正好是一个两党制国家希望看到的局面。

这种通过挖掘数据建立个人档案预测个人行为的微目标锁定技术经不断发展，已完全符合我们所说的数学杀伤性武器的所有特征。这种技术影响范围广，不透明，且不负责任。在该技术的掩护下，政客们得以更顺利地在不同的人面前展示不同的自己。

对个体选民进行评分也会削弱民主。这一评分系统让极少数的选民变得极为重要，而让其余的选民沦为配角。确实，从总统选举使用的模型来看，我们似乎处于一个被大大压缩了的国家。在写作本书的

时候，所有“高价值”的选民都居住在佛罗里达州、俄亥俄州、内华达州等摇摆州的少数几个郡，而这些郡当中也只有一小部分选民还在摇摆。我现在可以告诉你，如果说我们所讨论的许多数学杀伤性武器，如掠夺性广告、警力调度模型，损害的大多数是贫困阶级的利益，那么利用微目标锁定技术创建的选举模型损害的就是所有阶层的利益。不管你是来自曼哈顿还是来自旧金山，是贫穷还是富有，这些选民都被剥夺了事实上的选举权。（当然，那些真正的富豪仍然可以通过政治献金弥补损失。）

整个政治系统（通过金钱投入、关注、讨好等一系列活动）锁定特定的目标选民，就像一朵花追随太阳，而其余的选民则几乎被完全忽视了（除了能够提供资金的人）。模型已经预测到了我们大多数人的投票决定，且认为不值得投入任何资金来改变。注

一个恶性循环产生了。不被重视的选民会对国家的政治系统更加幻灭。胜利者了解游戏规则，他们知道内幕，而我们当中的绝大多数人只能接收经过市场测试的残羹冷炙。

此外，还存在另一种不平衡。预期会参与投票的人如果因为某种原因错过一轮选举，则这些人在下一轮选举中会得到特别关注。他们参与投票的可能性仍然很高。但是预期不会参与投票的大部分人都被忽视了。选举系统致力于寻找花最少的钱就能使其转换阵营的选民，希望所花费的每一分钱都能得到最大化的回报。而那些不投票者往往意味着更高的转化成本。这一动态系统刺激特定人群保持投票活跃度，同时确保余下的大多数继续保持低水平的政治参与度。

对于数学杀伤性武器来说，这种危害是常有的事，但正是这些模型也可以被用来造福人类。微目标锁定的目的可以不是控制他人，而是把他人排在接受帮助的队列里。例如，在某次市长竞选中，一个微目标锁定系统可能将一些选民归入了某个组别，因为他们都对难以承

受的高昂租金表示过愤怒。但是，如果市长候选人知道这些选民对房租有看法，那为何不用同样的技术定位那些可以从经济适用房项目中获益的人，并帮助他们争取这项福利呢？

正如大多数数学杀伤性武器一样，核心问题主要在于模型的目标。把目标由压榨大众变更为帮助大众，数学杀伤性武器的危险性就解除了，甚至可以反过来变成一种正面力量。

---

1. 类似地，购物网站更倾向于给还未登录的用户发放优惠券。这提醒我们定期清理浏览器缓存是很有必要的。

2. 就国家层面而言，废除总统选举团制度可以在很大程度上解决这个问题。极少数选民的高价值，来自选举中的赢家通吃算法，即最终的支持率是以州计而非以实际选票数计。这部分极少数的高价值选民，其地位就像经济领域站在金字塔尖的1%的特权阶级一样。这1%的特权阶级用自己的钱操控微目标锁定模型保障那1%的高价值选民的投票行为符合自己的利益。而如果废除了总统选举团制度，则每张选票就具有了同等的价值。这能让我们向更全面的民主更进一步。

结论



快速穿过虚拟的一生，我们经历了中学、大学、法庭和职场，甚至投票厅。一路上，我们见证了数学杀伤性武器的破坏力。数学杀伤性武器承诺效率和公平，却扭曲了高等教育，推高了债务，助长了大规模监禁，在每一个人生的重大关头打击穷人，破坏民主。合理的应对方式似乎就是一个一个解除这些武器。

问题是，这些数学杀伤性武器会互相滋养、固化。穷人更可能信用差，居住在犯罪高发区，周围的人也都是穷人。一旦数学杀伤性武器吃透了相关数据，它们就会向穷人的世界倾倒次级贷款或者营利性大学的掠夺式广告。治安系统会派更多的警力监管并逮捕他们，对他们处以更长的刑期。这些数据又会流入其他的数学杀伤性武器，从而使后者将他们判定为高风险人群或者易掌控群体，妨碍他们找到工作，同时推高他们的按揭、车贷以及你能想象得到的各方面保险的利率。穷人的信用等级因此进一步下降，一个完美的死亡漩涡就此诞生。在数学杀伤性武器的世界里，贫穷越发有害并且摆脱代价巨大。

同样的数学杀伤性武器既能压榨穷人，又能把社会富有阶层圈养在它们通过营销广告塑造的世界里。数学杀伤性武器把富人送去阿鲁巴岛度假，让他们竞争沃顿商学院的入学名额。而很多富人则感觉到这个世界似乎越来越智能便捷了。营销模型向他们推送意大利熏火腿和基安蒂红葡萄酒的促销信息，在亚马逊金牌会员服务的主页上为他们推荐新上映的大电影，或者引导富人探寻某个处于曾经的贫民街区的咖啡馆。这种微目标锁定模型使得富人们无法看到同样的模型是如何摧毁穷人生活的，而这些悲剧可能就发生在距他们咫尺之遥的地方。

美国的国家格言是“合众为一”，但数学杀伤性武器颠倒了这一等式。数学杀伤性武器暗箱操作，把民众整体进行分组归类，在幕后对我们的远亲近邻施加伤害，而且是极为广泛的伤害：一个单亲母亲因为频繁变更的工作时间表而不能为孩子及时安排好托儿所；一个努力奋斗的年轻人因为未通过职场人格测试而在申请小时工时被拒；少数族裔的穷小伙被警察无故拦截、粗暴对待，又被当地警局给予警告处分；一个来自贫困地区的加油站服务员不得不负担高额的保险账单。这是一场无声之战，穷人首当其冲，中产阶级也难逃一劫。大多数受害者经济实力欠缺，缺乏律师资源，也没有资金充足的政治组织为他们而战。结果是，穷人遭受了大范围的伤害，而上层阶级则称之为不可避免的牺牲。

我们不能指望自由市场纠正这些错误。为了解释其中的原因，我们可以把数学杀伤性武器与恐同症做一个类比，后者是我们社会一直致力于消除的又一个苦难的根源。

1996年9月，时任总统比尔·克林顿在连任竞选的两个月前签署了《婚姻保护法》。该法案明确定义婚姻是一男一女的合法结合，该法案的提出主要是为了争取俄亥俄州、佛罗里达州等几个保守州的选票。

仅一周之后，科技巨头公司IBM就宣布为公司职员中的同性伴侣提供医疗保险方面的福利。你可能会好奇，为什么该巨头公司会站在一位被公认为言行激进的美国总统的对立面为自己招惹争议呢?

答案和利益有关。1996年，互联网淘金热刚刚兴起，IBM正与甲骨文公司、微软公司、惠普公司以及众多初创科技公司争夺人才。这些公司大多数都已经为同性伴侣提供了特定福利，并且吸引了众多同性恋人才。IBM不能错失这个机会。“就商业竞争而言，这对我们有利。”IBM发言人在上述福利政策发表后对《商务周刊》表示。

如果把IBM和其他公司的人力资源策略看作算法，我们就会发现，嵌入编码的偏见已经存在几十年了。员工福利均等行动促进了公平。自此以后，性少数群体为很多领域的进步做出了显著的贡献。当然，这种进步不平衡。许多男同、女同和跨性别者仍然在遭受偏见、暴力和数学杀伤性武器的伤害，这种伤害在穷人和少数族裔中的性少数群体身上体现得尤其明显。不过，写作本书时，全世界最有价值的企业之一苹果公司，其首席执行官蒂姆·库克就是一个公开的男同性恋者。如果他愿意的话，他完全可以和他的男性伴侣结婚，这是宪法赋予他的正当权力。

现在我们已经知道，企业可以果断纠正其人力资源算法中的错误，那么，为什么企业不能以类似的态度和方法修正给整个社会造成严重危害的数学杀伤性武器呢?

不幸的是，这两者之间区别巨大。性少数者的权利保障事业的发展很大程度上受益于市场的力量。企业渴望雇用受教育水平高且逐步进入主流视野的性少数人才，所以企业优化了它们的人力模型以吸引性少数人才。但是企业这么做主要是出于利益的考虑，公平大多数时候只是副产品。与此同时，各行各业也开始争夺性少数群体中的富有的消费者，为他们提供游轮、特别优惠以及同性恋或跨性别题材的电

视节目。虽然这种包容性无疑会招致狭隘的保守人士的不满，但更重要的是，这一做法能为它们带来巨大的利润。

而拆除数学杀伤性武器并不总能带来这么明显的回报。虽然更高水平的公平和正义肯定有利于提升社会的整体利益，但单个企业无法从中获得实际的收益。其实，大多数企业认为数学杀伤性武器是一种非常高效的工具。很多行业的商业模型，如营利性大学和发薪日贷款，都是基于数学杀伤性武器而建的。当一个软件程序成功锁定那些极度需要金钱，愿意支付18%的月利率的人时，那些大肆敛财的公司就会认为软件运转良好，不存在问题。

当然，受害者的感受完全相反。但是大部分受害者都是穷人——小时工、失业人员，一辈子携带低信用评级的人，囚犯当然更是无能为力。在这个金钱可以买到影响力的社会，数学杀伤性武器的受害者几乎没有话语权。他们中的大多数都在事实上被剥夺了公民权。事实上，穷人常常被认为应该对自己的贫穷、糟糕的学校教育以及所在社区的混乱治安负有责任。这就是为什么很少有政治家会去费力制定反贫困政策。对他们而言，贫穷更像是一种疾病，他们通常采取的做法是隔离它，防止这种疾病传播到中产阶级。因此，我们需要思考的是，在现代生活中我们对贫困负有怎样的责任，以及数学杀伤性武器是如何加剧了恶性循环的。

但穷人也并不是数学杀伤性武器的唯一受害者——远非如此。我们已经看到，恶劣的数学模型能把合格的应聘者列入黑名单，克扣那些不符合公司所谓的理想健康指标的员工的薪酬。这些数学杀伤性武器对中产阶级造成了同样大的冲击。富人阶层也难逃一劫，盯上他们的是政治模型还有大学入学申请的评估模型。他们不得不像我们其他人一样疯狂地工作，以让自己达到无情的数学杀伤性武器设定的录取标准，而这一武器真正做的只是破坏高等教育的公平性。

需要特别指出的是，这些还都是早期的情形。发薪日贷款机构及其剥削者同类起初将目标对准了穷人和移民。这是理所当然的，因为穷人是最容易下手的目标，他们获取信息的渠道更少，有更多的人正处于绝望当中。但创造惊人利润空间的数学杀伤性武器不太可能长期在小范围内运作，这不符合市场的规律。它们会进化、传播，寻找新的机会。我们已经看到这种现象了，比如主流银行大规模投资借贷俱乐部等P2P贷款业务。简而言之，数学杀伤性武器已经开始将目标对准我们所有人了。而它们还将继续繁衍，播下不公正的种子，直到我们采取措施制止它们。

由贪婪或偏见导致的不公正一直发生在我们身边。你可能会争辩说，数学杀伤性武器并不比过去的人类更糟糕。毕竟在以往，信贷员或招聘经理可能会例行公事一般地排除整个黑人或女性获得抵押贷款或者一份工作的机会。许多人会说，即使是最糟糕的数学模型应该也没有那么糟糕。

但是，人类的决策虽然经常有缺陷，却也有一个主要的优点，即人类的决策是可以改善的。随着一步步地学习和适应，我们会改变，我们的处理方式也会改变。相比之下，自动化系统不会随着时间的推移而改变，除非开发者对系统做出改变。如果20世纪60年代就出现了大学申请的大数据模型，我们可能到现在还会有很多女性上不了大学，因为用于训练模型的数据都来自成功的男性。如果博物馆在那个年代将当时流行的伟大艺术思想进行编码并据此建立评价模型，我们可能到现在仍然只能看到白人男性艺术家的作品，因为当时由富人赞助进行艺术创作的都是白人男性。不用说，阿拉巴马大学的足球队也将仍然是全白人配置。

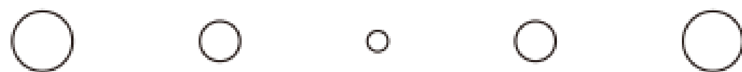
大数据程序只能将过去编入代码，而不会创造未来。创造未来需要道德想象力，而想象力只有人类才有。我们必须明确地将更好的价

值观嵌入我们的算法代码中，创造符合我们的道德准则的大数据模型。有时，这就意味着要重视公平，牺牲利润。

在某种意义上，我们的社会正在与一场新的工业革命做斗争。我们可以从上一次工业革命中吸取一些经验。20世纪是一个各行各业都取得了巨大进步的时代。人们得以在家里用电力照明，用煤取暖。现代铁路的兴建让人们品尝到了从遥远的大陆运来的肉类、蔬菜和罐头食品。许多人都感受到，生活正变得越来越美好。

然而，这种进步有一个可怕的阴暗面——进步的动力来源于被极度剥削的工人，其中甚至有很多是儿童。由于缺乏健康或安全方面的法规，煤矿成了死亡陷阱。仅在1907年一年，就有3242名矿工遇难。肉类加工员每天要在肮脏的环境下工作12~15个小时，并频繁地接触有毒的产品。肉类加工商Armour and Co.向美国军队运送了大量的烂牛肉罐头，使用硼酸掩盖罐头的臭味。与此同时，贪婪的垄断者主导着铁路、能源公司和公用事业公司，不断抬高消费税的税率，这相当于是在对整体国民经济征税。

显然，自由市场无法控制这些剥削行为。因此，直到艾达·塔贝尔和厄普顿·辛克莱等记者揭露了工厂中的种种问题之后，政府才开始介入。政府建立了食品安全条款和卫生检查制度，并取缔了童工。随着工会的兴起以及维护工人的法律的颁布，社会开始实行八小时工作制以及周末休息制度。因为竞争对手也必须遵守同样的规则，这些新标准保护了那些不想剥削工人或拒绝销售腐烂食品的公司。虽然工厂做生意的成本毫无疑问地增加了，但是整个社会因此受益。现在，我们当中恐怕没有人还愿意回到原来那个时代。



我们如何规范这些逐渐渗透继而操纵着我们整个生活的数学模型呢？我建议从建模者自己开始做起。数据科学家应该像医生一样遵守希波克拉底氏誓言，尽可能防止或避免对模型可能的误用和误解。2008年经济危机爆发之后，伊曼纽尔·德曼和保罗·威尔莫特两名金融工程师起草了这个誓言。内容如下：

> 我将牢记我并未创造世界，也不会让世界来满足我的方程式；
>
> 虽然我将大胆使用模型来估算价值，但不会过分倚重这一数学分析；
>
> 我将永远不会为了追求模型的简洁而牺牲现实的复杂，除非我能够对这样做的原因给出一个合理的解释；
>
> 我也不会向使用我所创建的模型的人们夸大模型的精准性。相反，我将明确说明模型中的假设条件和模型忽略的因素；
>
> 我明白我的工作可能会对社会和经济造成巨大的影响，其中的许多影响将超出我的认知范畴。

这是我们的数学杀伤性武器消除行动的一个很好的哲学基础。但坚定的价值观和自我约束只对谨慎的人有效。更重要的是，这一改编后的希波克拉底氏誓言忽略了数据科学家在寻求他们的老板所需求的具体答案时面对的实际压力。要想消灭数学杀伤性武器，我们不仅仅要在数据行业确立最佳范例，还要推行相关的法律条例。要想实现这一目标，我们必须重新评估模型的成功标准。

如今，一个模型的成功与否往往以利润、效率或者违约率等，为标准这几乎都是一些可以计算的标准。但是，我们真正应该计算的是什么呢？看看下面这个例子。当人们在搜索引擎上搜索食品券信息的时候，他们常常会遭遇中间商的广告插入，比如亚利桑那州坦佩市的“找到家庭资源”（FindFamilyResources）网站。此类网站看似官

方，且提供政府官方申请表页面的链接，但是，这些网站也为掠夺式广告商（包括营利性大学）收集申请人的姓名和电邮地址信息。此类网站通过给人们提供不必要的服务，将许多人纳入掠夺式广告商的模型锁定范围，而后者将疯狂地推送给他们超过他们消费能力的服务，网站则因此赚取巨额的潜在顾客开发费用。

这就是一次成功的模型应用吗？这要看你计算什么了。站在谷歌的角度，用户每点击一次广告就会产生25美分、50美分甚至几美元的收入，因此，这的确是成功的。顾客开发商自然也跟着赚了钱。因此看起来，这一系统正在高效运转，商业马车的车轮正在快速转动。

但是，站在社会的角度，穷人只是搜索了一下政府服务就被锁定了，很多穷人因此深陷虚假承诺和高利率贷款的泥潭。甚至从经济的角度来看，严格来说，这也是一种系统流失。人们需要食品券这一事实首先就表明了市场经济的萧条。政府企图利用大众缴纳的税款弥补这种萧条带来的损失，并寄希望于食品券能让申请者实现自给自足。但是，顾客开发商迫使穷人接触到了于他们毫无帮助的交易，很多学生因此欠债更多，以致对政府援助更加依赖。数学杀伤性武器为搜索引擎、顾客开发商、市场营销商创造了巨额收入，但其实质上是国家整体经济的血吸虫。

针对数学杀伤性武器建立的监管系统必须衡量这些潜在成本，同时还要体现更多的非数值价值。其他领域的监管法规已经将这些因素考虑在内了。虽然经济学家试图去计算雾霾、农田径流或者斑点猫头鹰的灭绝等成本，但是数字无法真正传达这些东西的价值。对于数学模型来说，公平和公共利益也是如此，其价值难以用数字体现。公平和公共利益是仅存在于人脑中的概念，无法量化。而且模型开发者往往也懒得思考这个问题，认为它太难了。但是，我们必须要在建立模型时考虑到这些价值，哪怕为此要牺牲效率。举个例子，你可以设计一个模型，其目的是确保选民群体或者消费者群体能够涵盖有各种种

族身份的人或各种收入水平的人，或者设计一个模型，以找到某些地区的人购买某种服务的价格是平均水平的两倍的情况。这些模型给出的近似值可能是很粗略的，尤其是在模型开发最开始的时候，但是这些近似值至关重要。数学模型应该是我们的工具，而不应该成为我们的主人。

成绩差距、大规模监禁以及选民对政治参与的冷漠是美国全国性的重大问题，不是自由市场或者数学算法能够解决的。所以，我们首先应该停止我们的科技乌托邦幻想，即无根据地寄希望于用算法和科技解决一切问题。要想让算法更好地服务于人类，我们必须承认算法不是全能的。

要想解除数学杀伤性武器，我们还需要衡量这种武器的影响范围，进行算法审查。在深入钻研程序代码之前，第一步要做的是做研究。我们首先应该把数学杀伤性武器看作一个黑盒，它吸收数据，吐出结论——某人有再次犯罪的中等风险，某人有73%的可能性给共和党候选人投票，某位教师评分排在最后的10%。通过研究这些输出结果，我们就能大致拼凑出模型的潜在假设，为模型的公平性打分。

有时候，有些数学杀伤性武器在最初被创建出来时很明显只是一个原始工具，用于把复杂的问题简单化，协助管理者更容易地开除一部分雇员或者选择特定消费者提供打折优惠。比如说，纽约公立学校使用的教师评估增值模型就是一个统计学上闹剧，前一年给蒂姆·克利福德的教学表现评了6分，第二年得分又飙升至96分。如果你将教师们每一年的得分绘制在一张图表里，则数据点分布的随机性可能和房间里的氢原子差不多。那些公立学校的大部分数学系的学生在15分钟之内就能够研究明白这些数据，然后自信地得出结论：该系统评分毫无意义。毕竟，优秀的教师往往一直以来都很优秀。跟垒球比赛中的替补球手不一样，对于教师而言，很少发生发挥超常的赛季紧跟着频频

失误的赛季这种情况。（还有一点不一样的是，教师的表现不适合量化分析。）

教师评估增值模型这种落后的模型没什么可修改的，唯一的应对方式就是丢弃这种不公平的模型，至少在下一个十年、二十年内不要再想着用数值模型来评估教师的教学水平了。教师的教学表现对于数学模型来说是一个太过复杂的概念，可用来训练模型的数据都是一些粗糙的替代变量。这种模型还没有完善到能够对我们信任的人做出重要评判的程度。教师考核需要考虑各种细微之处和具体情境，即使是在大数据时代，这仍然是一个很有挑战性的问题。

当然，不管是校长还是其他领导层，分析者都应该考虑很多数据，包括学生的测验成绩。此外，模型还应该具有积极的反馈回路。积极的反馈回路能够为建模者、数据科学家（以及自动化系统）提供反馈信息，用以对模型进行校正。在教师评估这一案例中，我们要做的就是询问老师和学生认为评分是否合理，以及他们是否理解和接受模型的前提，如果答案是否定的，那么应该如何改善模型。只有当我们的模型是一个拥有积极反馈回路的生态系统时，我们才有可能利用数据改善教育。在此之前，评估模型只是一种惩罚性的工具。

大数据支持者可能会很快指出，人类的大脑也在运行着内部模型，这些内部模型也往往内嵌着偏见或者利己主义。所以，内部模型的输出结果（在教师评估案例中就是对教师的评价）也必须要接受公平性方面的审查。而且审查方式要经过仔细的设计，并接受人类的测试，然后才可以推行自动化。与此同时，数学家可以设计模型帮助教师评估他们的教学效果然后加以改善。

对于其他模型的审查则更加复杂。比如说，美国很多州的法官在判刑前要参考再犯模型的评分。在该案例中，这一新兴的技术让我们既可以回溯从前，也可以预测未来。在接受了这一数学杀伤性武器给出的风险评估结果之后，法官的量刑模式是否会发生改变？毫无疑

问，在尚未有此模型之前，许多法官已经在其头脑中运行着同样令人不安的模型了，他们对贫穷的囚犯和少数族裔囚犯的处罚比对其他人的更严重。在有些情况下，可以想见，再犯模型可能会减少这种带有偏见的判断。但在大部分时候，再犯模型不会起到这种效果。在积累了足够多的数据之后，法官的判刑模式将变得清晰起来，然后，我们就可以评估这一数学杀伤性武器的优点和倾向性了。

如果我们发现（其实已经有研究表明）再犯模型内嵌着偏见假设，更倾向于惩罚穷人，那么我们现在就应该研究一下输入数据。在这个案例中，再犯模型包含大量的以“物以类聚，人以群分”为主要假设的替代变量。再犯模型依据一个人的朋友圈、职业以及信用等级等在法庭上不被承认的细节信息预测这个人将来的行为。而为了保证公平，解决方式就是丢弃这些方面的数据。

很多人可能会说，等一下，我们真的要为了追求公平而牺牲模型的精确度吗？简化算法真的是必要的吗？

在某些案例中，答案是肯定的。如果承认法律面前人人平等，或者作为选民的大众应该被平等对待，我们就不能允许模型把我们分为不同的群体进行区别对待。(注)亚马逊和网飞等公司会把付费客户细分，最大限度地从消费者身上榨取利润。而这样的算法不可能带来正义或者民主。

审查算法的行动已经在进行中了。例如，在普林斯顿大学，研究人员已经启动了“网络透明和问责项目”。他们设计了软件机器人，在互联网上伪装成富人、穷人、男性、女性以及患有精神疾病的人。通过研究这些机器人得到的反馈，这些研究员得以发现搜索引擎、招聘平台等自动化系统携带的偏见。卡耐基–梅隆大学和麻省理工学院等知名学府也有研究者在从事这一领域的研究。

审查行动得到学术支持至关重要。毕竟，要想监管数学杀伤性武器，我们需要有能力建立它们的专业人才。他们的研究工具能够复制数学杀伤性武器的规模化影响，并检索到足以展示出模型中的不平衡和不公正的数据集。他们也能够发起众包活动，让大众向他们提供自己收到的来自广告商或者政客的宣传信息的详情。如此，他们就可以阐明总统竞选中的微目标锁定技术的应用方法和策略。

当然，并不是所有模型都是恶性的。举例来说，2012年的总统选举结束之后，美国网络新闻机构ProPublica创立了“短信机器”（Message Machine），短信机器借助众包的方式逆向解码奥巴马竞选团队的定向政治广告所使用的模型。研究结果显示，不同的群体会听到不同的名人对奥巴马总统给出的赞美性言论，每一个名人锁定一个特定的观众群。这还不能算是确凿的证据，但至少，通过大众提供信息，消除竞选背后的模型的神秘感，短信机器有效地减少了（哪怕一点点的）谣传和猜测。这是件好事。

如果你认为数学模型是数字经济的引擎（当然在很多方面确实如此），那么，这些审查者所做的就是为我们揭开内幕，告诉我们数学模型是如何运作的。这是关键性的一步，在此基础之上，我们就可以为数学杀伤性武器这种强大的引擎配备方向盘——还有刹车了。

然而，审查者经常面临来自互联网巨头的阻挠，尽管互联网巨头公司建设的平台是我们利用数据的最便捷渠道。比如，谷歌禁止研究者制造大量虚假的个人简历去探索搜索引擎的偏见行为。[注]如果谷歌搜索算法确实内嵌偏见假设，那么公司肯定更愿意保守这个秘密，这样一来，谷歌就可以向外人隐藏算法的内幕和偏见了。但是谷歌的员工，他们也和我们一样存在证实性偏差，更容易看到他们想看到的东西。他们可能不会问出最尖锐的问题。而如果他们发现了算法的不公正，而这种不公正被用于增加谷歌的利润，这就会导致一场不愉快的谈话，这些也肯定是谷歌公司不想让公众知道的。所以，这些互联网

巨头有充足的商业理由支持保密条款。但随着公众对数学杀伤性武器了解得越来越多，并开始向这些机构问责时，谷歌将别无他法，只好选择信息公开，这就是我的希望。

脸书也是如此。脸书平台严格的实名制政策严重限制了外界人员的研究。实名制政策有许多好处，尤其体现在促使用户对他们发布的消息负责上。但脸书也必须对我们所有人负责，也就是向更多的数据审查者开放其平台。

当然，政府也要扮演一个更强大的监管角色，就像其在面对第一次工业革命时的剥削和悲剧时所做的那样。政府的第一步行动应该是修订相关法律条文，然后严格执法。

正如我们在信用评分系统那一章中所讨论的，美国民权法案《公平信用报告法》（FCRA）和《平等信用机会法》（ECOA）的推行旨在确保信用评分的公平性。《公平信用报告法》确保消费者可以看到某个数据档案对信用评分的影响，并且有权纠正档案中包含的任何错误，而《平等信用机会法》则禁止在信用评分中纳入种族或性别因素。

这些法规并不完美，迫切需要更新。消费者的抱怨经常被忽视，而且没有任何法规明确阻止信用评分公司把地理位置这一种族因素的替代变量纳入评估模型。尽管如此，这些法规还是开了一个好头。首先，我们应该要求提高透明度。当信用评分被任何机构用于评判或审查我们时，我们每个人都有权得到预先的警示。我们每个人都有权获得用于计算信用评分的信息。如果信息有错误，我们有权利质疑评估结果并纠正错误信息。

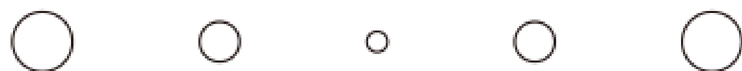
其次，这些法规的监管范围应该扩展到新型信贷公司，比如借贷俱乐部，此类公司利用最新的电子评分系统预测消费者的贷款违约风险。这些新型信贷公司不应该被获准进行暗箱操作。

《美国残疾人法案》也需要更新。该法案保护身体有缺陷的人在工作中不受歧视。该法案目前禁止将体检作为招聘筛选的一项参考。但法案需要把大数据人格测试、健康评分和声誉评分的危害性考虑在内，这些操作也都是违法的，不该被允许。目前正在讨论的一种修正是，将《美国残疾人法案》的保护延伸至对未来身体状况的“预测”结果。也就是说，如果一个基因组分析报告显示一个人患乳腺癌或阿尔茨海默病的风险很高，这个人也不应该被剥夺工作机会。

《医疗保险可携性和责任法案》也应该扩大对我们施予保护的范围。该法案保护我们的医疗信息，目的是防止我们的医疗信息数据被雇主、健康应用程序以及其他的大数据公司随意地收集和利用。任何由中间商收集的与健康有关的数据，包括在谷歌搜索医疗服务信息的访问记录数据，都必须得到保护。

如果想要更进一步，我们可以参考欧洲模式，即规定收集任何数据都必须经过用户的批准，用户具有选择权。欧洲模式还禁止将数据重新用于其他目的。“事前同意”由于常常被用户忽视，因此作用不大。但是，“不可重复使用”条款的约束力则非常强大：该条款将销售用户数据定义为非法行为。这也有效避免了数据落入数据中间商之手，后者往往利用手中掌握的用户数字档案支持有害的电子评分系统以及总统竞选活动中的微目标锁定行动。假定欧洲数据中间商会遵守法律，那么在“不可重复使用”的法规下，它们就会受到更多的限制。

最后一点是，那些对我们的生活有重大影响的模型，包括信用评分系统和各类电子评分系统，都应该对公众公开。理想情况下，我们应该可以像操作手机应用程序一样管理这些评分系统。例如，在某个月，某个女性消费者可能预算很紧张，那么她可以使用这种应用程序来比较未付的电话费账单和电费账单哪一个对她的信用评分影响更大，以及较低的系统评分会对她的购车计划产生多大的影响。制作此类手机应用程序的技术已经存在，只是我们没有这方面的意愿。



2013年的一个夏日，我乘地铁到曼哈顿南端，步行到纽约市政厅对面的一幢行政大楼。我的兴趣是建立和数学杀伤性武器作用相反的数学模型帮助社会。因此，我与纽约市住房和公共服务部门的数据分析小组签约，在该组作为无薪实习生工作。仅纽约市内，无家可归的人就已经增长到6.4万，其中包括2.2万名儿童。我的工作是要建立一个模型，用于预测一个无住房家庭对收容系统的依赖时间，然后为每个这样的家庭提供适合他们的服务。该模型的目的是给无家可归之人提供家庭生活所需，并为他们找到一个长期住所。

我在这个团队的任务在很大程度上相当于建立一个再犯模型。我和建立LSI–R模型的分析师很像，我感兴趣的是把人们再次推进庇护所的因素以及帮助他们得到稳定住房的因素。然而，与LSI–R模型不同的是，我们的团队致力于利用这些发现帮助受害者，减少无家可归、生活绝望的人。我们的模型的目标是维护公共利益。

在另一个与之相关的项目中，一位研究人员发现了一种非常强的相关性，这一相关性为我们的项目提供了一个解决方案。一些无住房的家庭往往会在某一天从收容所里默默消失，然后再也不会回来。这是因为他们受益于美国联邦经济适用住房计划，即“第8条”（Section 8），拿到了住房补贴。这没什么值得惊讶的。如果你为无住房家庭提供经济适用房，那么他们中的大多数也不会选择继续住在街道或脏乱的收容所了。

然而，这个结论让当时的市长迈克尔·布隆伯格及其政府感到了尴尬。在大张旗鼓的宣传下，为了让贫困家庭摆脱对“第8条”的依赖，纽约市政府创立了一个名为“优势”（Advantage）的新系统，该系统将住房补贴的发放时限限制在3年。政府认为，设定补助期限能够迫使穷人努力挣钱，自力更生。事实证明，政府的想法过于乐观了。而与此同

时，纽约市蓬勃发展的房地产市场极大地抬高了租金，这让穷人转向自力更生变得更加不可能。于是，没有补助金的家庭又回到了收容所。

研究人员的这一发现很不受欢迎。为了与高级政府官员会面，我们的小组准备了一个关于纽约无家可归者的幻灯片，其中一张幻灯片展示的是关于无家可归者再次回到收容所可能性的统计数据和“第8条”法案的有效性。展示结束后，我们团队和政府官员展开了一段非常尴尬而简短的对话。有人要求删掉这张幻灯片。最后，政党路线占了上风。

虽然大数据技术在使用得当时能够提供重要的参考意见，但这些参考意见中也有很多是具有破坏性的。毕竟，大数据的目的是寻找人类肉眼看不见的模式。数据科学家面对的挑战是理解他们猛烈抨击的大数据生态系统，而且不仅要找出问题所在，还要给出可能的解决方案。一个简单的工作流数据分析的结果可能会显示有5个职员是多余的。但是，如果数据团队引入专家，他们可能会发掘出该分析模型的更有建设性的优化版本，也就是说，优化后的模型可以为这5个人找到更适合他们的工作，帮助他们找到胜任这些工作所需要的培训。从某种意义而言，数据科学家的工作就是在你挖掘得不够充分的时候，继续深入挖掘数据。

当我审视整个数据经济时，我发现了大量新出现的数学模型，有些模型可被用作有益模型，有些模型，如果不被滥用的话，甚至有潜力成为伟大的模型。举个例子，米拉·伯恩斯坦是一个“奴隶侦探”，她是哈佛大学的数学博士，她创建了一个模型，用于扫描庞大的产业供应链，比如那些组装手机、运动鞋或越野车的工厂，目的是发现强迫劳动的迹象。该模型是她为一家名为“自由世界生产”（Made in a Free World）的非营利组织建立的，目标是帮助企业发现并根除其产生供应链中的剥削成分。企业迫切希望自己免受剥削成分的影响，有可能是

因为它们反对剥削，当然更可能是因为怕和剥削发生关联，破坏它们的品牌形象。

伯恩斯坦从多处收集数据，包括从联合国处获得的交易数据，奴役最盛行地区的统计数据，以及关于上万个工业产品的组装流程的详细信息等。她把所有这些信息都纳入模型中，训练该模型为某地区的一个给定产品就其存在强迫劳动现象的可能性进行评分。伯恩斯坦在接受《连线》杂志的采访时表示："我的假设是，用户会与供应商联系，让对方'告诉我我买的这台计算机的零部件都是从哪来的"。就像许多负责任的模型一样，这一奴役探测器不会做出过多干涉，它仅仅指向可疑的地方，问题的确定和解决最后还是由人类完成的。毫无疑问，公司会发现有些存在可疑之处的供应商并没有违规操作。（每一个模型都会有一定的误报率。）而这些信息会反馈给"自由世界生产"的模型，伯恩斯坦则可以据此校正模型。

还有一个以维护公共利益为目标的模型，出现在社会福利领域。这也是一个预测模型，用于准确锁定最有可能存在虐待儿童情况的家庭。该模型由美国东南部一个名叫Eckerd的儿童和家庭服务机构开发，于2013年在佛罗里达州的希尔斯伯勒县启动运行。在此前的两年里，该地区有9名儿童死于虐待，其中包括一名被扔出车窗的婴儿。该机构的训练数据库中囊括了1500个虐待儿童的案例，包括虐待至死的案例。机构工作者发现了许多指向虐待儿童的线索，比如住在家里的男朋友，家庭成员有过吸毒或家庭暴力的记录，父母中的某一方曾在寄养家庭长大。

如果这是一个锁定潜在犯罪分子的模型，你可能会立即发现该模型是多么的不公平。在寄养家庭长大或者家里有一个未婚伴侣不应该成为被怀疑的理由。而且，这种模型更有可能锁定贫困人群，从而忽视可能发生在富有家庭的虐待儿童案。

但是，该模型的目标不是惩罚父母，而是为可能需要帮助的孩子提供帮助，如此，这个潜在的数学杀伤性武器就转变为一个良性模型。该模型被用于统筹资源为有虐待儿童风险的家庭提供帮助。根据《波士顿环球报》的报道，在该模型启动后的两年里，希尔斯伯勒县没再发生虐待儿童至死的案件。

有理由相信，类似这样的良性模型将在未来几年里大量出现，评估我们患骨质疏松或中风的风险，帮助学习有困难的学生学习微积分，甚至预测哪些人最有可能遭受重大生活变故。就像我们讨论过的一些数学杀伤性武器一样，其中一些模型，其出发点是好的，但要保证这些模型的良性运转，还必须使其公开透明，对公众披露模型使用的输入数据以及输出结果。此外，模型必须接受审查。毕竟，这些模型可是强大的机器，我们必须看好它们。

数据不会消逝，计算机也不会，数学更不会。预测模型日益成为我们的必备工具，我们利用这些工具经营各种机构，配置资源，管理我们的生活。但是，本书想要阐明的是，这些模型的建立不仅仅基于数据本身，也基于我们关注或忽视哪些数据的选择。我们的选择不仅关乎物流、利润和效率，从根本上来说，这是一个道德问题。

如果我们回避对数学杀伤性武器的探索，把它们当作一种中立的力量、不可避免的趋势，就像天气或潮汐，我们就等于是放弃了我们的责任。我们已经发现，数学杀伤性武器视我们为机器零件、无足轻重的抗议者，以不公平为养料。我们必须团结起来监管这些数学杀伤性武器，驯服它们并解除它们的武装。我希望，在我们学会如何为这个数据时代注入公平和问责之后，将来的人们在回忆起数学杀伤性武器时，会把它当作这场新革命的早期遗产，就像一个世纪前的致命煤矿一样。数学比数学杀伤性武器的价值大得多，民主也是如此。

---

1. 你可能会认为，对于模型公平性的审查可能会要求模型去除其中的与种族特质相关的变量。但如果我们要考察一个数学杀伤性武器在制造种族偏见方面的危害性，我们就需要拿数据来说话。而目前，绝大部分数学杀伤性武器都不会直接追踪用户的种族信息。在很多情况下，这一做法都是违法的。但在另一些情况中，比如申请房屋按揭贷款，种族身份是申请表中必填的一项。因此相比较于汽车贷款申请，房屋按揭贷款的申请筛选更容易发生种族偏见。因此，为考察模型在制造种族偏见方面的危害，我们应该考察房屋按揭贷款申请过程使用的筛选模型，以数据说明这种模型造成种族不平等的严重性和广泛性。然后，我们可以将这一结果公开发表，进行道德伦理方面的辩论，并提出相应的补救措施。虽然这么说，但种族身份其实是一种社会建构，即便你想针对这一议题进行针对性的研究，这一概念所涉及的范畴也并没有那么容易把握，这一点，任何一个多种族混血儿都能向你证明。
2. 一些谷歌员工跟我说，谷歌已经表示出主动消除其搜索算法中的偏见假设的意愿。而我的回应是，希望谷歌能向外部的研究者开放其数据平台。

致谢



感谢我的丈夫和孩子们的鼎力支持。同时也感谢约翰·强森、史蒂夫·沃尔德曼、稻田珍姬、贝基·杰夫、亚伦·艾布拉姆斯、朱莉·斯蒂尔、凯伦·彭斯、马特·拉曼蒂亚、潘玛莎、丽莎·拉德克利夫、路易斯·丹尼尔和梅利莎·比尔斯基。还要感谢罗拉·斯加奥斯费德、阿曼达·库克、艾玛·贝利、乔丹·艾伦伯格、史蒂芬·贝克、杰伊·曼德尔、山姆·可颂–比纳纳夫和厄尼·戴维斯，没有你们的帮助就不会有这本书。